N° d'ordre : 03 ISAL 0068
Année 2003

Thèse

Consultation assistée par ordinateur de la documentation en Sciences Humaines

Considérations épistémologiques, solutions opératoires et applications à l'archéologie

Présentée devant
L'Institut National des Sciences Appliquées de Lyon
Pour obtenir
Le grade de docteur
École doctorale
Informatique et Information pour la Société
Spécialité
Documents Multimédia, Images et Systèmes d'Information Communicants

Par
Aurélien Bénel (Ingénieur)

Soutenue le
Le 12 décembre 2003
Devant le jury interdisciplinaire suivant :

Roland Etienne	Professeur, Université Paris I	Président
Jocelyne Nanard	Professeur, Université Montpellier II	Rapporteur
François Rastier	Directeur de Recherche, INaLF (CNRS) Paris	Rapporteur
Jean-Marie Pinon	Professeur, INSA de Lyon	Directeur
Andrea Iacovella	Ingénieur de Recherche, Ecole française d'Athènes	Co-directeur
Sylvie Calabretto	Habilitée à Diriger les Recherches, INSA de Lyon	Co-directrice

Thèse préparée à l'École française d'Athènes [EFA] et au Laboratoire d'InfoRmatique en Images et Systèmes d'information [LIRIS]

Résumé

Dans les Sciences Humaines, comme dans bon nombre de domaines d'expertise (Médecine, Ingénierie, Droit, Physique…), le document constitue le principal outil de travail et, à ce jour, ni les bases de données ni les systèmes experts n'ont réussi à le supplanter. Dans ces systèmes, en effet, l'usager serait contraint d'insérer ses données dans un modèle fixe et consensuel (schéma de la base, ontologie). Or, dans les domaines d'expertise, chacun apporte son modèle personnel vis-à-vis d'un corpus documentaire de référence. De plus, la confrontation des points de vue stimule une évolution constante des modèles.

Né d'une dynamique interdisciplinaire (Archéologie, Informatique…), le projet *Porphyre* se propose d'instrumenter ce travail d'expertise. Il est fondé sur l'idée qu'une telle pratique « laisse des traces » : sélections de fragments, documents ajoutés au corpus, organisation de la bibliographie, notes de lecture, etc. A condition que ces traces soient « captées », leur auteur peut désormais les utiliser pour accéder au corpus, ainsi que les partager avec d'autres experts. En somme, puisque le corpus se structure de pair avec le savoir de l'expert au fil de ses changements de problématique et de point de vue, nous proposons un atelier de manipulation de corpus documentaires afin d'assister un travail humain de construction de sens.

Mots-clés : Bibliothèque numérique, Assistance à l'interprétation, Collaboration, Annotation, Hypermédia.

Abstract

In Humanities (as in Medicine, Engineering, Law, Physics...), neither database systems nor expert systems succeeded in supplanting documents. In fact, the expert cannot insert her data in a fixed and consensual model (database model, ontology). On the contrary, she needs to build her own hypothetic model regarding a corpus of reference documents. Moreover, the confrontation of different points of view stimulates a constant evolution of models.

Born from an interdisciplinary dynamics (Archaeology, Computer Science...), the *Porphyry Project* proposes a hypermedia solution. The main idea is that the human expertise "leaves traces": documents added to the corpus, highlighted fragments, organization of the bibliography, reading notes, etc. Provided that these traces are "collected", they can be used as a corpus structure for their author and her community. By offering a workshop for handling documentary corpora, we hope to assist humans in constructing the *meaning*.

Keywords: Digital Libraries, Interpretation Assistance, Collaboration, Annotation, Hypermedia.

Ecoles Doctorales et Diplômes d'Etudes Approfondies

habilités pour la période 1999-2003

ECOLES DOCTORALES n° code national	RESPONSABLE PRINCIPAL	CORRESPONDANT INSA	DEA INSA n° code national	RESPONSABLE DEA INSA
CHIMIE DE LYON **(Chimie, Procédés, Environnement)** **EDA206**	M. D. SINOU UCBL1 04.72.44.62.63 Sec 04.72.44.62.64 Fax 04.72.44.81.60	**M. R. GOURDON** 87.53 Sec 84.30 Fax 87.17	Chimie Inorganique **910643** Sciences et Stratégies Analytiques **910634** Sciences et Techniques du Déchet **910675**	 M. R. GOURDON Tél 87.53 Fax 87.17
ECONOMIE, ESPACE ET MODELISATION DES COMPORTEMENTS **(E^2MC)** **EDA417**	M.A. BONNAFOUS LYON 2 04.72.72.64.38 Sec 04.72.72.64.03 Fax 04.72.72.64.48	**Mme M. ZIMMERMANN** 60.91 Fax 87.96	Villes et Sociétés **911218** Dimensions Cognitives et Modélisation **992678**	Mme M. ZIMMERMANN Tél 60.91 Fax 87.96 M. L. FRECON Tél 82.39 Fax 85.18
ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE **(E.E.A.)** **EDA160**	**M. D. BARBIER** INSA DE LYON 85.47 Fax 60.82		Automatique Industrielle **910676** Dispositifs de l'Electronique Intégrée **910696** Génie Electrique de Lyon **910065** Images et Systèmes **992254**	M. M. BETEMPS Tél 85.59 Fax 85.35 M. D. BARBIER Tél 85.47 Fax 60.82 M. J.P. CHANTE Tél 87.26 Fax 85.30 Mme I. MAGNIN Tél 85.63 Fax 85.26
EVOLUTION, ECOSYSTEME, MICROBIOLOGIE , MODELISATION **(E2M2)** **EDA403**	**M. J.P FLANDROIS** UCBL1 04.78.86.31.50 Sec 04.78.86.31.52 Fax 04.78.86.31.49	**M. S. GRENIER** 79.88 Fax 85.34	Analyse et Modélisation des Systèmes Biologiques **910509**	M. S. GRENIER Tél 79.88 Fax 85.34
INFORMATIQUE ET INFORMATION POUR LA SOCIETE **(EDIIS)** **EDA 407**	**M. L. BRUNIE** INSA DE LYON 87.59 Fax 80.97		Documents Multimédia, Images et Systèmes d'Information Communicants **992774** Extraction des Connaissances à partir des Données **992099** Informatique et Systèmes Coopératifs pour l'Entreprise **950131**	M. A. FLORY Tél 84.66 Fax 85.97 M. J.F. BOULICAUT Tél 89.05 Fax 87.13 M. A. GUINET Tél 85.94 Fax 85.38
INTERDISCIPLINAIRE SCIENCES-SANTE **(EDISS)** **EDA205**	**M. A.J. COZZONE** UCBL1 04.72.72.26.72 Sec 04.72.72.26.75 Fax 04.72.72.26.01	**M. M. LAGARDE** 82.40 Fax 85.24	Biochimie **930032**	M. M. LAGARDE Tél 82.40 Fax 85.24
MATERIAUX DE LYON **UNIVERSITE LYON 1** **EDA 034**	**M. J. JOSEPH** ECL 04.72.18.62.44 Sec 04.72.18.62.51 Fax 04.72.18.60.90	**M. J.M. PELLETIER** 83.18 Fax 85.28	Génie des Matériaux : Microstructure, Comportement Mécanique, Durabilité **910527** Matériaux Polymères et Composites **910607** Matière Condensée, Surfaces et Interfaces **910577**	M. J.M.PELLETIER Tél 83.18 Fax 85.28 M. H. SAUTEREAU Tél 81.78 Fax 85.27 M. G. GUILLOT Tél 81.61 Fax 85.31
MATHEMATIQUES ET INFORMATIQUE FONDAMENTALE **(Math IF)** **EDA 409**	**M. F. WAGNER** UCBL1 04.72.43.27.86 Fax 04.72.43.00.35	**M. J. POUSIN** 88.36 Fax 85.29	Analyse Numérique, Equations aux dérivées partielles et Calcul Scientifique **910281**	M. G. BAYADA Tél 83.12 Fax 85.29
MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE **(MEGA)** **EDA162**	**M. F. SIDOROFF** ECL 04.72.18.61.56 Sec 04.72.18.61.60 Fax 04.78.64.71.45	**M. G.DALMAZ** 83.03 Fax 04.72.89.09.80	Acoustique **910016** Génie Civil **992610** Génie Mécanique **992111** Thermique et Energétique **910018**	M. J.L. GUYADER Tél 80.80 Fax 87.12 M. J.J.ROUX Tél 84.60 Fax 85.22 M. G. DALMAZ Tél 83.03 Fax 04.78.89.09.80 M. J. F. SACADURA Tél 81.53 Fax 88.11

En grisé : Les Ecoles doctorales et DEA dont l'INSA est établissement principal

INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

***Directeur* : STORCK A.**

***Professeurs* :**

AUDISIO S.	PHYSICOCHIMIE INDUSTRIELLE
BABOT D.	CONT. NON DESTR. PAR RAYONNEMENTS IONISANTS
BABOUX J.C.	GEMPPM***
BALLAND B.	PHYSIQUE DE LA MATIERE
BAPTISTE P.	PRODUCTIQUE ET INFORMATIQUE DES SYSTEMES MANUFACTURIERS
BARBIER D.	PHYSIQUE DE LA MATIERE
BASTIDE J.P.	LAEPSI****
BAYADA G.	MECANIQUE DES CONTACTS
BENADDA B.	LAEPSI****
BETEMPS M.	AUTOMATIQUE INDUSTRIELLE
BIENNIER F.	PRODUCTIQUE ET INFORMATIQUE DES SYSTEMES MANUFACTURIERS
BLANCHARD J.M.	LAEPSI****
BOISSON C.	VIBRATIONS-ACOUSTIQUE
BOIVIN M. (Prof. émérite)	MECANIQUE DES SOLIDES
BOTTA H.	UNITE DE RECHERCHE EN GENIE CIVIL - Développement Urbain
BOTTA-ZIMMERMANN M. (Mme)	UNITE DE RECHERCHE EN GENIE CIVIL - Développement Urbain
BOULAYE G. (Prof. émérite)	INFORMATIQUE
BOYER J.C.	MECANIQUE DES SOLIDES
BRAU J.	CENTRE DE THERMIQUE DE LYON - Thermique du bâtiment
BREMOND G.	PHYSIQUE DE LA MATIERE
BRISSAUD M.	GENIE ELECTRIQUE ET FERROELECTRICITE
BRUNET M.	MECANIQUE DES SOLIDES
BRUNIE L.	INGENIERIE DES SYSTEMES D'INFORMATION
BUREAU J.C.	CEGELY*
CAVAILLE J.Y.	GEMPPM***
CHANTE J.P.	CEGELY*- Composants de puissance et applications
CHOCAT B.	UNITE DE RECHERCHE EN GENIE CIVIL - Hydrologie urbaine
COMBESCURE A.	MECANIQUE DES CONTACTS
COUSIN M.	UNITE DE RECHERCHE EN GENIE CIVIL - Structures
DAUMAS F. (Mme)	CENTRE DE THERMIQUE DE LYON - Energétique et Thermique
DOUTHEAU A.	CHIMIE ORGANIQUE
DUFOUR R.	MECANIQUE DES STRUCTURES
DUPUY J.C.	PHYSIQUE DE LA MATIERE
EMPTOZ H.	RECONNAISSANCE DE FORMES ET VISION
ESNOUF C.	GEMPPM***
EYRAUD L. (Prof. émérite)	GENIE ELECTRIQUE ET FERROELECTRICITE
FANTOZZI G.	GEMPPM***
FAVREL J.	PRODUCTIQUE ET INFORMATIQUE DES SYSTEMES MANUFACTURIERS
FAYARD J.M.	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS
FAYET M.	MECANIQUE DES SOLIDES
FERRARIS-BESSO G.	MECANIQUE DES STRUCTURES
FLAMAND L.	MECANIQUE DES CONTACTS
FLORY A.	INGENIERIE DES SYSTEMES D'INFORMATIONS
FOUGERES R.	GEMPPM***
FOUQUET F.	GEMPPM***
FRECON L.	REGROUPEMENT DES ENSEIGNANTS CHERCHEURS ISOLES
GERARD J.F.	INGENIERIE DES MATERIAUX POLYMERES
GERMAIN P.	LAEPSI****
GIMENEZ G.	CREATIS**
GOBIN P.F. (Prof. émérite)	GEMPPM***
GONNARD P.	GENIE ELECTRIQUE ET FERROELECTRICITE
GONTRAND M.	PHYSIQUE DE LA MATIERE
GOUTTE R. (Prof. émérite)	CREATIS**
GOUJON L.	GEMPPM***
GOURDON R.	LAEPSI****.
GRANGE G.	GENIE ELECTRIQUE ET FERROELECTRICITE
GUENIN G.	GEMPPM***
GUICHARDANT M.	BIOCHIMIE ET PHARMACOLOGIE
GUILLOT G.	PHYSIQUE DE LA MATIERE
GUINET A.	PRODUCTIQUE ET INFORMATIQUE DES SYSTEMES MANUFACTURIERS
GUYADER J.L.	VIBRATIONS-ACOUSTIQUE
GUYOMAR D.	GENIE ELECTRIQUE ET FERROELECTRICITE
HEIBIG A.	MATHEMATIQUE APPLIQUEES DE LYON
JACQUET-RICHARDET G.	MECANIQUE DES STRUCTURES
JAYET Y.	GEMPPM***
JOLION J.M.	RECONNAISSANCE DE FORMES ET VISION
JULLIEN J.F.	UNITE DE RECHERCHE EN GENIE CIVIL - Structures
JUTARD A. (Prof. émérite)	AUTOMATIQUE INDUSTRIELLE
KASTNER R.	UNITE DE RECHERCHE EN GENIE CIVIL - Géotechnique
KOULOUMDJIAN J.	INGENIERIE DES SYSTEMES D'INFORMATION
LAGARDE M.	BIOCHIMIE ET PHARMACOLOGIE
LALANNE M. (Prof. émérite)	MECANIQUE DES STRUCTURES
LALLEMAND A.	CENTRE DE THERMIQUE DE LYON - Energétique et thermique
LALLEMAND M. (Mme)	CENTRE DE THERMIQUE DE LYON - Energétique et thermique
LAUGIER A.	PHYSIQUE DE LA MATIERE

Mai 2003

LAUGIER C.	BIOCHIMIE ET PHARMACOLOGIE
LAURINI R.	INFORMATIQUE EN IMAGE ET SYSTEMES D'INFORMATION
LEJEUNE P.	UNITE MICROBIOLOGIE ET GENETIQUE
LUBRECHT A.	MECANIQUE DES CONTACTS
MASSARD N.	INTERACTION COLLABORATIVE TELEFORMATION TELEACTIVITE
MAZILLE H.	PHYSICOCHIMIE INDUSTRIELLE
MERLE P.	GEMPPM***
MERLIN J.	GEMPPM***
MIGNOTTE A. (Mle)	INGENIERIE, INFORMATIQUE INDUSTRIELLE
MILLET J.P.	PHYSICOCHIMIE INDUSTRIELLE
MIRAMOND M.	UNITE DE RECHERCHE EN GENIE CIVIL - Hydrologie urbaine
MOREL R.	MECANIQUE DES FLUIDES ET D'ACOUSTIQUES
MOSZKOWICZ P.	LAEPSI****
NARDON P. (Prof. émérite)	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS
NIEL E.	AUTOMATIQUE INDUSTRIELLE
NORTIER P.	DREP
ODET C.	CREATIS**
OTTERBEIN M. (Prof. émérite)	LAEPSI****
PARIZET E.	VIBRATIONS-ACOUSTIQUE
PASCAULT J.P.	INGENIERIE DES MATERIAUX POLYMERES
PAVIC G.	VIBRATIONS-ACOUSTIQUE
PELLETIER J.M.	GEMPPM***
PERA J.	UNITE DE RECHERCHE EN GENIE CIVIL - Matériaux
PERRIAT P.	GEMPPM***
PERRIN J.	INTERACTION COLLABORATIVE TELEFORMATION TELEACTIVITE
PINARD P. (Prof. émérite)	PHYSIQUE DE LA MATIERE
PINON J.M.	INGENIERIE DES SYSTEMES D'INFORMATION
PONCET A.	PHYSIQUE DE LA MATIERE
POUSIN J.	MODELISATION MATHEMATIQUE ET CALCUL SCIENTIFIQUE
PREVOT P.	INTERACTION COLLABORATIVE TELEFORMATION TELEACTIVITE
PROST R.	CREATIS**
RAYNAUD M.	CENTRE DE THERMIQUE DE LYON - Transferts Interfaces et Matériaux
REDARCE H.	AUTOMATIQUE INDUSTRIELLE
RETIF J-M.	CEGELY*
REYNOUARD J.M.	UNITE DE RECHERCHE EN GENIE CIVIL - Structures
RIGAL J.F.	MECANIQUE DES SOLIDES
RIEUTORD E. (Prof. émérite)	MECANIQUE DES FLUIDES
ROBERT-BAUDOUY J. (Mme) (Prof. émérite)	GENETIQUE MOLECULAIRE DES MICROORGANISMES
ROUBY D.	GEMPPM***
ROUX J.J.	CENTRE DE THERMIQUE DE LYON – Thermique de l'Habitat
RUBEL P.	INGENIERIE DES SYSTEMES D'INFORMATION
SACADURA J.F.	CENTRE DE THERMIQUE DE LYON - Transferts Interfaces et Matériaux
SAUTEREAU H.	INGENIERIE DES MATERIAUX POLYMERES
SCAVARDA S.	AUTOMATIQUE INDUSTRIELLE
SOUIFI A.	PHYSIQUE DE LA MATIERE
SOUROUILLE J.L.	INGENIERIE INFORMATIQUE INDUSTRIELLE
THOMASSET D.	AUTOMATIQUE INDUSTRIELLE
THUDEROZ C.	ESCHIL – Equipe Sciences Humaines de l'Insa de Lyon
UBEDA S.	CENTRE D'INNOV. EN TELECOM ET INTEGRATION DE SERVICES
VELEX P.	MECANIQUE DES CONTACTS
VIGIER G.	GEMPPM***
VINCENT A.	GEMPPM***
VRAY D.	CREATIS**
VUILLERMOZ P.L. (Prof. émérite)	PHYSIQUE DE LA MATIERE

Directeurs de recherche C.N.R.S. **:**

BERTHIER Y.	MECANIQUE DES CONTACTS
CONDEMINE G.	UNITE MICROBIOLOGIE ET GENETIQUE
COTTE-PATAT N. (Mme)	UNITE MICROBIOLOGIE ET GENETIQUE
ESCUDIE D. (Mme)	CENTRE DE THERMIQUE DE LYON
FRANCIOSI P.	GEMPPM***
MANDRAND M.A. (Mme)	UNITE MICROBIOLOGIE ET GENETIQUE
POUSIN G.	BIOLOGIE ET PHARMACOLOGIE
ROCHE A.	INGENIERIE DES MATERIAUX POLYMERES
SEGUELA A.	GEMPPM***
VERGNE P.	LaMcos

Directeurs de recherche I.N.R.A. **:**

FEBVAY G.	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS
GRENIER S.	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS
RAHBE Y.	BIOLOGIE FONCTIONNELLE, INSECTES ET INTERACTIONS

Directeurs de recherche I.N.S.E.R.M. **:**

PRIGENT A.F. (Mme)	BIOLOGIE ET PHARMACOLOGIE
MAGNIN I. (Mme)	CREATIS**

** **CEGELY*** *CENTRE DE GENIE ELECTRIQUE DE LYON*
*** **CREATIS*** *CENTRE DE RECHERCHE ET D'APPLICATIONS EN TRAITEMENT DE L'IMAGE ET DU SIGNAL*
**** **GEMPPM*** *GROUPE D'ETUDE METALLURGIE PHYSIQUE ET PHYSIQUE DES MATERIAUX*
***** **LAEPSI*** *LABORATOIRE D'ANALYSE ENVIRONNEMENTALE DES PROCEDES ET SYSTEMES INDUSTRIELS*

Sommaire

Introduction 1
1ère partie : État de l'art 3
Chapitre 1. Préliminaires méthodologiques 5
Chapitre 2. Statut des modèles informatiques : Le cas des ontologies 17
Chapitre 3. Explication et compréhension en Sciences Humaines 31
Chapitre 4. Instrumentation de l'intertextualité et de l'intersubjectivité 39
2ème partie : Modèle proposé 53
Chapitre 5. Traces et publication 55
Chapitre 6. Contenus documentaires 65
Chapitre 7. Réseaux de description 73
Chapitre 8. Parcours de lecture 95
3ème partie : Études de cas 105
Chapitre 9. La *Chronique des fouilles* du BCH 107
Chapitre 10. CEFAEL : Collections de l'Ecole française d'Athènes en ligne 115
Chapitre 11. La nécropole occidentale de Mégara Hyblaea 125
Chapitre 12. Perspectives 135
Conclusion 143
Bibliographie 147
Table des matières 177
Annexes 187
Annexe A : Syntaxe utilisée pour les spécifications algébriques I
Annexe B : Historique du prototype III
Annexe C : Pourquoi diffuser *Porphyre* sous Licence GPL ? V
Annexe D : Manuel d'utilisation du client *Porphyre* VII
FOLIO ADMINISTRATIF X

Et vive les ornithorynques !

Remerciements

- Aux institutions qui ont apporté leur soutien au projet :
  Le Ministère délégué de la Recherche et de l'Enseignement Supérieur,
  Le Programme « Société de l'Information » du CNRS,
  Le Réseau Thématique Pluridisciplinaire « Document » du CNRS,
- Aux directions et aux personnels des structures qui ont hébergé nos recherches : l'Ecole française d'Athènes [EFA] et le Laboratoire LIRIS (INSA de Lyon),
- A Roland Etienne (EFA), Andrea Iacovella (EFA), Sylvie Calabretto (LIRIS) et Jean-Marie Pinon (LIRIS) sans qui ces travaux n'auraient jamais eu lieu,
- A Jocelyne Nanard, François Rastier qui m'ont fait l'honneur d'accepter d'être membres de mon jury,
- A tous les stagiaires ou doctorants ayant participé à la conception, au développement et à la diffusion de *Porphyre* : Franck Eyraud, Thomas Buisson, Mehdi Lababidi, Laurent Pinel, Elodie Tasia, Rodolphe Vatré, Jocelyn Viallon, Olivier Chadenat, Olivier Martin, Rémi Huynh, Michel Nux, Tiphaine Accary, Régine Tribollet, Julien Gossa, Baptiste Meurant, Guillaume Deshors, Caroline Djambian et Florent Ruard-Dumaine,
- A ceux qui ont eu l'extrême gentillesse de relire ma prose,
- A tous ceux qui, à un moment ou à un autre, ont manifesté de l'intérêt pour mes recherches.

Introduction

Les Sciences Humaines font partie de ces domaines « d'expertise » (Recherche, Médecine, Ingénierie, Droit…) dans lesquelles le document constitue le principal outil de travail. Cet outil, à ce jour, ni les bases de données ni les systèmes experts n'ont réussi à le supplanter. Le corpus documentaire est, en effet, la référence ultime du discours de l'expert.

La présente thèse se propose d'explorer la question de l'instrumentation informatique de la consultation par les experts de ces corpus. En fait, le corpus se structurant de pair avec le savoir de l'expert au fil de ses changements de problématique et de point de vue, il devient envisageable d'assister un travail humain de construction de sens, en offrant « simplement », un atelier de manipulation de corpus documentaires.

Dans la première partie de cette thèse, après avoir resitué historiquement et institutionnellement notre étude, nous tenterons de dresser un état de cette question de l'assistance à la consultation de corpus pour experts. Nous verrons comment des considérations épistémologiques, portant sur les Sciences en général et les Sciences Humaines en particulier, nous permettent de nous orienter dans un catalogue qui autrement aurait été pléthorique.

Fort de notre critique des travaux existants, nous présenterons dans une deuxième partie un nouveau modèle appelé « *Porphyre* ». Nous verrons comment ce modèle opératoire[1] est fondé sur l'idée qu'une pratique interprétative « laisse des traces », traces qui pourront plus tard être retrouvées et susciter le débat entre experts.

Une troisième partie, dédiée à des études de cas en archéologie, nous permettra de mettre le doigt sur les difficultés rencontrées lors de l'usage de notre modèle. Pour y faire face, des solutions de natures diverses (modèle, usage, formation) seront proposées. Enfin, dans un dernier chapitre, nous verrons comment notre équipe prévoit, à tra-

[1] Au sens « d'exécutable par une machine ».

vers de nouveaux projets et de nouvelles thèses, de donner à ces travaux des perspectives à moyen et plus long terme.

1ère partie : État de l’art

Chapitre 1. Préliminaires méthodologiques

A beginning is a very delicate time.
Frank Herbert, *Dune*, 1965.

La thèse que nous défendons ici est celle d'un informaticien qui a tenté de se laisser « déplacer » par d'autres disciplines au sein du réseau interdisciplinaire constitué et animé par Andrea Iacovella depuis cinq ans.

Pour autant, est-il correct de dire qu'il s'agit d'une « thèse interdisciplinaire » ? Assurément non, une telle expression serait un oxymoron. En effet, rien n'est plus disciplinaire qu'une thèse. Notre thèse ne peut donc qu'être profondément marquée par les méthodes et le discours de la discipline informatique. Autrement dit, si l'on comparait le réseau interdisciplinaire à un polyèdre, notre travail constituerait juste une facette, parmi celles des archéologues, historiens, architectes, historiographes…

Ce travail prendra donc tout son sens une fois accompagné de ceux des autres acteurs du réseau. En attendant, il est toutefois possible, dans ce chapitre préliminaire, de retracer la dynamique du réseau interdisciplinaire, et de replacer ainsi cette thèse dans sa situation d'écriture.

Dans une première partie, nous verrons comment le réseau doit son émergence à une problématique issue des sciences historiques. Ensuite, dans une deuxième partie, nous présenterons l'approche adoptée par le réseau, une approche à la fois interdisciplinaire et opératoire. Par la suite, dans une troisième partie, nous dresserons une liste des difficultés rencontrées puis évitées par le réseau. Enfin, dans une dernière partie, nous tenterons de définir l'objet d'étude commun aux différentes disciplines représentées.

1. Une problématique issue des sciences historiques

Le point de départ de ce réseau réside dans un questionnement théorique sur la création du sens dans les Sciences Historiques. En particulier, Andrea Iacovella a eu l'occasion à plusieurs reprises d'identifier comme une impasse méthodologique le cou-

rant de l'archéologie défendant la définition normative d'une typologie des vestiges (ou l'utilisation de thesaurus pour l'interrogation de « banques de données »).

> À la différence du spécialiste des Sciences de l'Information et de la Communication, occupé à affecter des significations non ambivalentes et non ambiguës en regard des corpus documentaires, l'historien manipule un objet qui déborde de sens de toutes parts. Dans un index, un catalogue, un titre d'ouvrage ou dans un corps de texte, le mot « Auschwitz » ne se laisse pas résumer à de l'information. Écrit en toutes lettres ou représenté par son équivalent binaire, Auschwitz reste Auschwitz. Dans le seul fait de prononcer un mot comme celui-là, il y a le risque d'une banalisation qui effraie l'esprit ; le travail de l'historien, n'est pas tant de produire des catégories bien huilées, mais d'élaborer un langage qui permettra à des mots comme « Auschwitz » de se maintenir en toute dignité dans une phrase et de contribuer par là à donner tout son sens à une communauté du dialogue. [IacovellaEtAl03]

Comme le montre la gravité de l'exemple, la question du nécessaire renouvellement des méthodes dans les sciences historiques est relayée par les grands débats de société actuels. Comment se fait-il par exemple que les méthodes de l'Histoire n'aient pu endiguer des dérives telles que celles du négationnisme portant sur la Shoah ? Même s'il est reconnu aujourd'hui un écart entre Méthode et Vérité, peut-on pour autant dire n'importe quoi du moment que l'on adopte le discours de l'érudit ? Puisque l'activité scientifique (quelle que soit la discipline) est incapable de juger du caractère éthique d'une théorie, n'est-il pas nécessaire de faire appel au politique (au sens large) ? Sans recourir forcément à des « comités d'éthique », la solution se trouve peut-être dans un nouveau type de « vulgarisation » qui ôterait à ce mot son caractère péjoratif, autrement dit, dans une meilleure interaction entre le débat scientifique et le débat civique.

On aurait pu croire que le problème serait réglé par l'avènement de la « Société de l'Information ». Au contraire, sur la Toile, le problème est accentué par :

- l'explosion du nombre de pages personnelles, reflet d'un certain relativisme postmoderne (prônant que « tout est vrai »),
- la « confiscation » des revues scientifiques par des monopoles éditoriaux imposant peu à peu des modèles économiques rendant leur accès inabordable pour l'amateur « éclairé ».

Ainsi, le but scientifique du réseau vient à la rencontre d'un but plus politique, celui de passer de la « Société de l'information » à une nouvelle dynamique à inventer : la « Société de la Connaissance ».

2. Une approche à la fois interdisciplinaire et opératoire

La double nature (scientifique/éthique) de la problématique nécessitait de « tresser » ensemble deux « fils » : celui de la Recherche et celui du Service (c'est-à-dire de l'action). Ce n'est donc nullement un hasard si cette dynamique a pu voir le jour sous la direction d'Andrea Iacovella, à la fois chercheur en historiographie et responsable du Service Informatique de l'Ecole française d'Athènes (EFA).

La première étape de cette dynamique consista à évaluer la faisabilité de la mise en ligne de la *Chronique des fouilles* publiée par l'EFA. Dans cette perspective, une collaboration fut entreprise entre l'EFA et le Département informatique de l'INSA de Lyon sous la forme d'un certain nombre de stages [Komninidis95, Larcher97, Montocchio97, Benel98]. Il en ressortit que, si les technologies étaient relativement mûres pour la mise en ligne de collections, il subsistait en informatique un « verrou » concernant la gestion de leur sémantique. Andrea Iacovella décida, avec le soutien de Roland Etienne alors directeur de l'EFA, d'associer au projet de mise en ligne des collections, un projet de recherche interdisciplinaire ayant pour but de faire tomber ce « verrou ».

a. Mise en ligne des collections scientifiques de l'EFA

D'abord consacré à la seule *Chronique des fouilles*, le projet de mise en ligne fut rapidement généralisé au *Bulletin de Correspondance Hellénique* [BCH], puis à l'ensemble des collections publiées par l'EFA depuis ses 150 ans d'existence, soit près de 250 000 pages en 570 volumes. Les *Collections de l'Ecole française en ligne* [CEFAEL] constituent ainsi une bibliothèque numérique à la fois exhaustive et publique.

En 2001, CEFAEL fit l'objet d'un prototypage complet, portant sur la chaîne de numérisation et la diffusion sur la Toile du BCH. Le déroulement de cette expérimentation fut rendu possible par la signature d'une convention entre l'EFA et la Maison de

l'Orient et de la Méditerranée [MOM]. Les résultats de cette expérimentation reçurent le franc soutien du Ministère de la Recherche. Ce dernier accorda alors sur le « Fonds National pour la Science » le budget nécessaire pour passer en phase de production. Par la suite, furent étudiées les conditions de l'hébergement de CEFAEL au Centre Informatique National de l'Enseignement Supérieur [CINES].

A mi-parcours du projet, furent organisées des « Journées Bibliothèques Numériques[2] » afin d'apporter la contribution de l'équipe au débat sur les bibliothèques numériques en France, en particulier en termes de mise en place de modèles de production et de diffusion. A cet effet, un guide pratique[3] fut distribué à tous les participants.

Après achèvement de la numérisation, saisie des sommaires, finalisation du modèle des données et transfert sur les serveurs du CINES, le portail CEFAEL[4] fut inauguré le 9 avril 2003.

b. Recherches sur la gestion de la sémantique des collections scientifiques

Le processus que nous venons de présenter s'est déroulé de pair avec un travail de recherche sur le renouvellement des méthodes dans les disciplines historiques. Plus particulièrement, notre équipe s'est attachée à trouver une solution interdisciplinaire au « verrou » de la discipline informatique évoqué plus haut.

Cette orientation interdisciplinaire s'est affirmée en particulier dans l'organisation d'une table ronde[5] sur le thème « Sémantique et Archéologie » rassem-

---

[2] Journées d'études des 29-30 mai 2002, organisées à Lyon par l'Ecole française d'Athènes, la Maison de l'Orient Méditerranéen et l'Université Lyon 2, avec le soutien de la Sous Direction des Bibliothèques (Ministère de l'Education Nationale).

[3] *Modèles opératoires de production et de diffusion des collections scientifiques dans les bibliothèques numériques* (expérimentés à partir de la numérisation des publications de l'Ecole française d'Athènes), Manuel publié dans le cadre des journée d'études « Bibliothèques Numériques », Lyon, 29-30 mai 2002.

[4] http://cefael.efa.gr

[5] Sémantique et Archéologie : Aspects expérimentaux (Renouvellements méthodologiques dans les bibliothèques numériques et les publications scientifiques), Table ronde interdisciplinaire, Athènes, 18-19 novembre 2000.

blant des intervenants issus de l'Archéologie, de l'Historiographie, de la Sémiotique, des Sciences Cognitives, et de l'Informatique. La dynamique entamée par cette table ronde s'est poursuivie par la création d'un réseau interdisciplinaire (EFA, LISI[6], MOM, GRACO[7]) se retrouvant périodiquement dans des séminaires de travail.

Les aspects méthodologiques du problème ne firent pas oublier la nécessité d'offrir des solutions opératoires. Aussi, une convention entre l'EFA et le LISI fut signée en 2000 pour la création d'un prototype informatique (appelé « Porphyre »). De même, le volet « recherche en informatique » fut encouragé par le lancement de notre thèse sous la codirection d'Andrea Iacovella pour l'EFA et de Jean-Marie Pinon (assisté de Sylvie Calabretto) pour le LISI.

Les travaux du réseau interdisciplinaire furent grandement soutenus par le CNRS et le Ministère de la Recherche. De 1999 à 2002, notre thèse fit l'objet d'une allocation de recherche. De plus, les départements SHS[8] et STIC[9] du CNRS, dans le cadre du Programme « Société de l'Information », nous accordèrent en 2001 un financement pour un projet sur la modélisation du temps dans les publications archéologiques mises en ligne[10] et en 2003 sur la sémiotique spatiale du document d'architecture en archéologie[11]. Il y a peu, le réseau a déposé au département STIC du CNRS une demande pour constituer une équipe projet[12].

---

[6] LISI : Laboratoire d'Ingénierie des Systèmes d'Information, Lyon. Aujourd'hui intégré au LIRIS (Laboratoire d'InfoRmatique en Images et Système d'Information).

[7] GRACO : Groupe de Recherche sur l'Antiquité Classique et Orientale, Toulouse.

[8] SHS : Sciences Humaines et Sociales.

[9] STIC : Sciences et Techniques de l'Information et de la Communication.

[10] « Assistance dans la gestion de ressources intertextuelles multiformes. Production et intégration interactives de parcours interprétatifs ».

[11] « Modélisation opératoire d'un système d'interrogation du document d'architecture en archéologie à base d'un langage visuel : application aux corpus numérisés et à la modélisation des connaissances ».

[12] « Du partage de corpus de documents structurés à la confrontation de points de vue ».

3. Cartographie des écueils

Si l'on peut se réjouir aujourd'hui du trajet parcouru par le réseau interdisciplinaire face à cette question de la gestion de la sémantique des collections scientifiques, il est cependant nécessaire de noter que tout cela ne s'est pas fait sans mal. L'interdisciplinarité a ceci de particulièrement délicat qu'y faire appel, c'est reconnaître qu'aucune discipline ne possède la réponse à la question posée [IacovellaEtAl03]. Elle se distingue en ceci de la pluridisciplinarité. Par exemple, lorsque l'archéologue demande au physicien de lui dater un vestige au radiocarbone, il s'agit de pluridisciplinarité. Au contraire, si les incohérences de cette mesure amènent le physicien à mettre en doute des hypothèses tacites de sa discipline, il s'agit d'un premier pas vers l'interdisciplinarité.

En l'absence d'un *manuel de l'interdisciplinarité*, nous avons été contraints de « naviguer à vue » entre les écueils. Certains de ces écueils sont peut-être encore présents dans cette thèse, cependant nous avons tenté à chaque « collision » de les « cartographier » et de les éviter par la suite. En espérant que ceci puisse bénéficier à d'autres projets interdisciplinaires, nous allons maintenant tenter de dresser une liste de ces écueils.

a. Création d'une nouvelle discipline

Le premier danger qui pèse sur l'interdisciplinarité consiste en l'affranchissement à l'égard des disciplines d'origine, en une fondation comme discipline scientifique autonome [IacovellaEtAl03]. En effet, quand une discipline se forme par fusion de deux autres (comme par exemple la biophysique à partir de la biologie et de la physique), on serait en droit de se demander si un débat entre deux disciplines ne serait pas plus stimulant pour la Science que la création d'un consensus local.

Conscient de ce danger, nous avons résolument choisi durant notre thèse de ne pas cantonner nos publications aux seuls colloques interdisciplinaires [BenelEtAl99, BenelEtCalabretto00, Benel00], mais de soumettre également des articles dans des colloques clairement identifiés comme appartenant à la discipline informatique [BenelE-

tAl00a, BenelEtAl00b, BenelEtAl01a, BenelEtAl02]. Ainsi nos pairs, en acceptant nos articles, ont pu nous signifier notre appartenance à la discipline informatique.

b. Instrumentalisation et réification des disciplines

Dans une première version de ce chapitre, nous avions analysé les buts poursuivis par l'EFA, ses flux d'information avec l'extérieur, son organisation interne, etc. En somme, nous avions appliqué naïvement les méthodes d'analyse de systèmes d'information. On nous a reproché, à raison, de réifier par là la discipline archéologique. En effet, dans le cas qui nous occupe, une véritable posture interdisciplinaire nécessite d'instrumenter l'archéologie sans pour autant l'instrumentaliser. La relation entre STIC et SHS que nous tentons de promouvoir dans cette thèse est la suivante :

- les SHS proposant des « grilles de lecture » aux théories des STIC,
- les STIC proposant des instruments aux pratiques en SHS.

Dans tous les cas, il est nécessaire, d'une part, que soit réciproque l'application de méthodes d'une discipline à l'autre, et d'autre part, que cette application soit matière à un questionnement théorique des disciplines.

c. Impérialisme d'une discipline

Au cours de la constitution du réseau, la collaboration avec certaines disciplines fut tellement difficile que le responsable du réseau dut prendre la difficile décision d'y mettre fin. Avec le recul, l'impossibilité de travailler avec ces disciplines s'expliquerait peut être par leur aspiration à devenir des « méta-disciplines », des censeurs des autres disciplines. En effet, comme nous l'évoquions plus haut, la condition *sine qua non* pour qu'un chercheur s'implique dans une démarche interdisciplinaire, c'est de reconnaître que quelque chose « échappe » à sa discipline.

d. Dissymétrie des « produits »

Aujourd'hui lorsque l'on parle du réseau interdisciplinaire, la plupart de nos interlocuteurs nous demandent si *Porphyre* fonctionne et si l'on peut leur en faire une démonstration. Il y a là un grave danger de rester rivé aux seuls aspects opératoires [Ia-

covellaEtAl03]. Reste dans l'ombre non seulement tout l'aspect théorique des recherches informatiques, mais aussi, ce qui est plus grave encore, la part des Sciences Humaines dans le projet.

Si la part de chacun dans le projet apparaît aujourd'hui de manière aussi faussée, c'est probablement que nous n'avons pas été suffisamment attentifs à respecter une certaine symétrie des produits[13] non seulement entre théorie et expérimentation[14], mais encore entre SHS et STIC. Ainsi, il pourrait être intéressant qu'à l'issue d'un cycle de séminaires chaque intervenant écrive quelques pages présentant son point de vue, disciplinaire, sur le débat qui a eu lieu. Ces synthèses, accessibles par tous, permettraient ainsi de donner une meilleure transparence de la dynamique interdisciplinaire qui se joue dans le réseau.

4. Définition de l'objet interdisciplinaire

Ayant toutes ces précautions à l'esprit, on pourrait nous reprocher d'avoir pris comme objet d'étude interdisciplinaire le verrou existant au sein des STIC concernant la gestion de la sémantique des collections scientifiques. Cependant, il est nécessaire de préciser que le seul fait de prendre un objet d'une discipline et de le livrer à l'étude des autres disciplines change cet objet.

> Lorsque nous disons, par exemple, que les *logiques temporelles* sont un thème de recherches interdisciplinaires associé au projet, ça ne signifie pas que par *logiques temporelles* on entend leur sens théorique érigé en dogme de référence […] et qui donnerait un statut majeur au discours de l'informaticien sur les autres disciplines. L'historien, l'archéologue, l'historiographe sont associés au projet comme « experts » du temps historique et à ce titre ils contribuent à un approfondissement de cette catégorie conceptuelle à l'intérieur des logiques temporelles. [IacovellaEtAl03]

[13] Rapports, articles scientifiques, prototypes, manuels, etc

[14] A titre d'illustration, *Porphyre*, avec la vingtaine de stagiaires qui ont participé à son développement, représente 18 000 lignes de code source, alors que cette thèse en représente environ 4 000.

Parce que la question de la consultation des documents concerne différentes communautés au sein de la discipline informatique, il devient nécessaire, avant d'aller plus loin, de voir comment ce « verrou » se décline d'une communauté à l'autre.

a. Recherche d'informations

La première communauté, celle de la « recherche d'information », a été popularisée par les moteurs de recherche sur la Toile (Altavista, Google…). Ses techniques consistent en général à indexer les documents à l'aide des mots du texte qui les discriminent le plus du corpus. Notons que d'un point de vue philosophique il n'est nullement question ici de « contenu »[15], seules étant considérées les propriétés statistiques de la « forme » (distribution des mots, des liens hypertextes, etc.).

Cette approche s'explique par l'histoire du concept d'information dans la discipline informatique. En 1948, lorsque Claude E. Shannon, ingénieur chez Bell, établit sa « théorie de l'information » [Shannon48], il quantifia l'information en fonction de la nouveauté d'un message par rapport aux messages reçus, c'est à dire comme ce qui est statistiquement improbable. Son but était d'obtenir un modèle permettant d'augmenter le rendement de la transmission à travers un réseau de communication. Son travail, d'ailleurs, a eu et continue d'avoir des applications spectaculaires en télécommunication et en compression de données.

Cependant, selon Bernard Miège (en Sciences de la Communication), l'application de cette théorie à d'autres domaines, même couronnée d'un certain succès, serait l'affaire d'un terrible quiproquo [Miege95]. Tout d'abord, on reproche souvent à la théorie de ne pas tenir compte de la composante sémantique du message. Plus précisément, le concept d'information peut s'opposer parfois à celui de signification, puisque, comme l'a expliqué Umberto Eco [Eco62], plus un message sera signifiant (non ambigu, redondant), moins il sera porteur d'information. D'autre part, le modèle est

[15] Même si paradoxalement on parle de « recherche par le contenu » en référence au « texte intégral ».

celui d'un canal de communication : ni l'émetteur ni le récepteur ne sont modélisés[16]. Les seules ambiguïtés considérées sont celles dues au « bruit » présent sur le canal. Enfin, le modèle requiert un ensemble fini de messages possibles, donné *a priori*, ce qui oblige en pratique à appliquer le modèle à des signaux simples voire simplistes (suites de caractères de longueur limitée).

b. Modélisation des connaissances

La seconde communauté est celle de la « modélisation des connaissances » dont les technologies les plus connues sont les systèmes de bases de données et de connaissances. Dans le domaine documentaire, sa méthode consiste à réaliser des modèles formels des besoins de l'usager et du contenu des documents. Un document est apparié à un besoin si l'on peut déduire à partir du modèle du premier le modèle du second. C'est le cas par exemple de l'approche booléenne utilisée dès les années 60 [Salton68] dans les catalogues informatisés des bibliothèques. Un document indexé par « philosophe ET mortel » sera apparié à la requête « philosophe OU grec ». De même, dans des systèmes plus récents, un document indexé par un graphe modélisant « Le philosophe Socrate est mortel » sera apparié au graphe requête modélisant « Un homme est mortel ».

Cette approche, participant au projet de « l'intelligence artificielle », est calquée sur la démonstration automatique de théorèmes. Elle consiste à démontrer des « buts » en appliquant successivement des règles de transformation à des « faits » de base.

Aujourd'hui, les plus vives critiques de l'approche viennent de la communauté elle-même. Ainsi, John F. Sowa, inventeur des célèbres « graphes conceptuels », dénonce-t-il l'idéologie dominante de sa communauté comme étant devenu une « nuisance obstructive » [Sowa01a]. Pour conforter son propos, il nous présente le parcours de Terry Winograd. Celui-ci écrivit en 1972 un livre intitulé « Comprendre le langage naturel » où il exposait le système SHRDLU capable de dialoguer avec un être humain au sujet de cubes empilés les uns sur les autres. En 1986, après avoir discuté avec des phé-

[16] Il est représentatif de voir qu'à un colloque de recherche d'information comme SIGIR, la « recherche d'information interactive » est souvent tenue à l'écart des conférences plénières et reléguée aux « ateliers ».

noménologues, il expliquait les faiblesses de sa première approche : l'oubli du contexte, du cadre de référence, du besoin de négociation et des inévitables exceptions. Des aspects que le courant philosophique à l'origine du domaine avait volontairement mis de côté. Convaincu que l'intelligence artificielle menait à une impasse, Terry Winograd travaille aujourd'hui dans le domaine des interactions homme-machine.

c. Hypertexte

La troisième communauté à considérer est celle de l'hypertexte (ou hypermédia). Chacun de nous a tendance à l'associer au « World Wide Web », c'est-à-dire à une version simplifiée inventée en 1989-90 par le CERN pour ses besoins propres [BernersLee89].

La notion, quant à elle, remonte aux années 1945 [Bush45][17]. Il s'agissait à l'époque de construire une machine (mécanique) permettant de « feuilleter » des microfilms. Le principe était en effet le suivant : associer à un *contenu* documentaire interprétable uniquement par l'homme, une *structure* (ou parcours) gérable par une machine.

Si l'approche a un succès retentissant, certains chercheurs remarquent néanmoins que de nombreux lecteurs d'hypertextes se trouvent désorientés, « perdus dans l'hyperespace ». En effet, contrairement à un texte, il est souvent difficile de remettre en situation un hypertexte, qu'il s'agisse de sa situation d'écriture (Qui a écrit quoi et quand ?) ou de lecture (Où suis-je ? D'où viens-je ? Où puis-je aller ?). En somme, bien souvent, l'hypertexte serait moins qu'un texte : il serait un « hypotexte » [Bachimont99b].

d. Quelle grille de lecture ?

Dans les trois communautés considérées, nous avons vu qu'il existait des critiques fondamentales portant non sur les techniques elles-mêmes, mais plus sur le cadre

[17] Le terme, lui, n'apparaît que vingt ans plus tard (THEODOR H. NELSON, « The Hypertext », in *Proceedings of the World Documentation Federation*, 1965).

philosophique dans lequel elles sont mises en œuvre. Il serait donc vain de s'en tenir à des typologies purement informatiques.

On mesure alors en quoi les débats menés au sein du réseau interdisciplinaire nous ont été indispensables. Néanmoins, ces apports n'apparaîtront pas en tant que tel puisqu'il nous était impossible d'appuyer notre argumentaire sur des échanges verbaux. Nous avons donc tenté de trouver des articles ou des monographies y faisant écho. Les prises de position des participants restant très avant-gardistes dans la discipline archéologique, nous avons été contraints à généraliser notre « état de l'art » aux Sciences Humaines voire même pour certains aspects aux Sciences en général. En termes purement méthodologiques, ceci revient juste à prendre plus de risques, étant donné que plus un discours scientifique prétend à l'universalité et plus il sera facile à réfuter…

Chapitre 2. Statut des modèles informatiques : Le cas des ontologies*

LE LOGICIEN, AU VIEUX MONSIEUR
Voici donc un syllogisme exemplaire. Le chat a quatre pattes. Isidore et Fricot ont chacun quatre pattes. Donc Isidore et Fricot sont chats.
LE VIEUX MONSIEUR, AU LOGICIEN
Mon chien aussi a quatre pattes.
LE LOGICIEN, AU VIEUX MONSIEUR
Alors c'est un chat […]
LE LOGICIEN, AU VIEUX MONSIEUR
Autre syllogisme : tous les chats sont mortels. Socrate est mortel. Donc Socrate est un chat.
LE VIEUX MONSIEUR
Et il a quatre pattes. C'est vrai, j'ai un chat qui s'appelle Socrate.
Ionesco, Rhinocéros.

Le discours des chercheurs en informatique au sujet des « ontologies » (en particulier dans le cadre du « Web sémantique »), nous semble révélateur d'une certaine indécision dans la communauté concernant le statut de leurs modèles.

L'une des définitions les plus acceptées de « l'ontologie informatique » est celle de Thomas R. Gruber [Gruber93] comme « spécification explicite d'une conceptualisation ». Il est donc question de langage et de concept, en d'autres mots, d'une théorie de la connaissance. Cependant, suivant les auteurs (voire chez un même auteur), le but des « ontologies » serait :

1. l'intelligence artificielle (« des machines comprennent des humains »),
2. l'interopérabilité (« des machines comprennent des machines »),
3. de faire sens pour l'usager (« des humains comprennent des humains via des machines »).

---

* Ce chapitre a fait l'objet d'une conférence dans le cadre d'une table-ronde sur le « Web sémantique » organisée pour le séminaire annuel de l'Institut Scientifique du Document Numérique [Benel02a]. De plus, certains éléments ont été présentés durant nos conférences :
- à la table-ronde « Sémantique et Archéologie » [Benel00],
- à la « European Conference on Digital Libraries » [BenelEtAl01a].

Or, nous pensons que ces trois buts proviennent de trois théories de la connaissance, issues de courants non seulement différents, mais surtout incompatibles.

Nous aimerions exhumer les postulats implicites (voire inconscients) qui déchirent la communauté. Dans ce but, nous proposons une petite étude de textes. Étant donné que dans les « sciences formelles », on hésite souvent à faire apparaître dans les articles de recherche des discussions plus philosophiques, notre corpus comprendra également des ouvrages de synthèse et des articles de vulgarisation.

Dans l'étude qui suit, nous adopterons une grille de lecture inspirée de la sémiotique[19], distinguant trois « mondes » : le monde empirique, le monde des idées et le monde du discours.

1. Valeur ontologique

a. Définition

Le mot « ontologique » est formé sur deux racines grecques :

- *onta*, les choses qui existent
- *logos*, le discours (raisonné).

On dira qu'un discours a une valeur « ontologique » (cf. Figure 2.1), lorsque celui-ci représente une idée qui elle-même représente le réel. En termes kantiens, on pourrait parler de raisonnement « *a posteriori* », c'est-à-dire postérieur à l'observation.

[19] Nos trois mondes s'inspirent en effet des trois facettes du signe : Référent/Signifié/Signifiant [Eco73]. Considérons le signe /cheval/. Son référent est l'ensemble des chevaux ayant existé, existant, et qui existeront. Son signifié est le concept de "cheval". Un concept peut exister en l'absence de référent (par exemple le concept de "licorne"). Son signifiant peut être un dessin, un mot écrit ou encore prononcé.

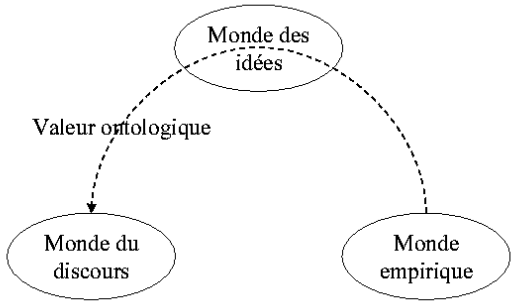


Figure 2.1 : Valeur ontologique du discours

Par exemple le mot « homme » désigne une foule hétéroclite d'individus différents mais dont l'appartenance doit être évidente pour quiconque en rencontrerait un. Aristote dit d'ailleurs dans le dernier chapitre des Seconds Analytiques [Leonhardt02] que lorsque l'on rencontre Callias, c'est l'homme en général que nous atteignons par la sensation et non Callias[20].

b. Le mythe de l'intelligence artificielle

Que signifierait un statut ontologique des modèles informatiques ? Ni plus ni moins que l'aboutissement du programme de l'intelligence artificielle. En effet, l'ordinateur, en opérant sur le discours, manipulerait directement des connaissances sur le monde réel.

Hubert Dreyfus [Dreyfus72] fut l'un des premiers à identifier l'intelligence artificielle comme étant un mythe[21], entraînant la fin du financement de la plupart des projets américains et européens[22]. Aussi, quand on sait la capacité remarquable d'un mythe à résister à l'oubli, on ne devrait pas s'étonner outre mesure de le voir resurgir, vingt ans après la crise, dans certaines orientations de recherche.

---

[20] Un bon contre-exemple de l'immédiateté de la classification est donné par l'ornithorynque (ovipare mais allaitant ses petits), cauchemar des taxinomistes qui, en désespoir de cause, l'ont contraint à rentrer dans la classe des mammifères.

[21] tout au moins dans la traduction française de son livre (le titre orignal étant un peu moins fort).

[22] Ce que la communauté lui reprocha durement [Pitrat84] mais qui ne faisait qu'anticiper la décision japonaise après l'échec des « ordinateurs de cinquième génération ».

Ainsi, même John F. Sowa (qui par ailleurs prend clairement un parti anti-positiviste) place ses ontologies informatiques dans la plus pure tradition de l'ontologie philosophique :

> Logic is pure form, and ontology provides the content that is expressed in that form. [Sowa00]

Ce à quoi, on pourrait facilement rétorquer :

> Comme le dit Blanché : « Un moule à raisonnement n'est pas un raisonnement, pas plus qu'un moule à gâteaux ne peut être mangé comme dessert ». [Borillo84, p.25]

Les promoteurs du « Web sémantique » [BernersLeeEtAl01], quant à eux, redécouvrent toutes les propriétés du langage utopique rêvé par le courant ontologique :

- un langage universel basé sur la logique : « Its unifying logical language will enable these concepts to be progressively linked into a universal web. »,
- un langage monosémique et non contextuel : « information is given well defined meaning »,
- un langage référentiel : « the Semantic Web will break out of the virtual realm and extend into our physical world. URIs[23] can point to anything, including physical entities »,
- enfin, un langage artificiel compréhensible par l'ordinateur : « machines become much better able to process and "understand" the data that they merely display at present. ».

L'exemple le plus célèbre est sans doute le projet *Cyc* [Wood00] contenant aujourd'hui un demi million de « faits ». Ce projet a nécessité le travail de 25 informaticiens pendant 20 ans. Ses coûts de développement (50 M$) furent financés par des multinationales ainsi que par la Défense Américaine (DoD). Nous retiendrons deux résultats de ce projet.

Le premier est une démonstration récente, dans les locaux du Pentagone, qui a réussi à associer par une suite de savantes inférences « anthrax » avec « Ben Laden ». En guise de commentaire, nous citerons :

> L'une des premières tâches est de discerner le statut théorique de ces activités [intelligence artificielle] et il n'est pas sûr que la seule façon d'y

[23] URI : Uniform Resource Identifier.

> parvenir, ni la meilleure, soit de parler de machines qui « jouent » aux échecs, qui « reconnaissent » des images, qui « comprennent » des textes… même si cette phraséologie est encore abondamment employée par les chercheurs eux-mêmes et si elle évoque assez bien, d'un point de vue externe, les fonctions que semblent remplir ces systèmes. Le malheur, c'est que de telles formulations sont doublement trompeuses : pour les profanes, qu'elles mystifient en accréditant le mythe des machines « pensantes » (« insuffler aux calculateurs numériques une intelligence à l'image de celle des hommes » !) ; et trop souvent des chercheurs qu'elles entraînent dans de faux challenges scientifiques (« battre » des joueurs de niveau plus élevé, « reconnaître » plus d'images, « comprendre » plus de phrases…) [Borillo84, p.15-16]

Le second est une remarque de Doug Lenat (le chef du projet) : « vampires are not real but in fictional settings they may be treated as real ». Nous dirions quant à nous qu'il existe des concepts comme « vampire » qui sans faire référence à des « choses qui sont » ont toutefois un certain pouvoir explicatif.

c. Courant philosophique

On peut faire remonter l'approche ontologique à Aristote [Aristote-300]. Par définition, il s'agit d'un discours raisonné sur l'Être. Dans une telle approche, le langage est considéré comme reflet exact du réel, à condition que le philosophe (aujourd'hui on dirait « le chercheur ») suive un certain nombre de règles. Tout d'abord, afin d'amorcer le raisonnement avec des principes indubitables, le philosophe doit traduire dans un langage artificiel (le langage catégorique) la connaissance évidente du réel qui lui est donnée (on serait tenté de parler de « révélation »). Ensuite, afin d'avancer dans le raisonnement, le philosophe doit user de syllogismes valides.

Il est important de noter que l'approche ontologique fait l'objet d'une autofondation. En effet, la méthode utilisée est elle-même considérée comme évidemment vraie, ontologique. Pour les disciples de cette approche, il n'y a aucun doute sur le bienfondé ni du langage catégorique comme apte à refléter le réel, ni sur le choix des syllogismes valides, ni sur la capacité du philosophe à connaître le réel et à le traduire en langage catégorique.

Cette approche a en fait traversé le temps, d'Aristote (IV° s. av. J.C) au Positivisme Logique (début du XX°s.), en passant par Descartes, Laplace, Comte et bien d'autres. Bien loin de s'affaiblir, elle s'est même radicalisée. En effet, Aristote se gardait bien d'appliquer sa méthode destinée à l'étude de la Nature (la physique) à d'autres domaines comme la métaphysique ou la politique. Au contraire le Positivisme Logique considère que la méthode ontologique est le modèle unique de la raison humaine.

2. Valeur logique

a. Définition

Le mot « logique » est formé sur la racine grecque *logos*, discours (raisonné). On dira qu'un discours a une valeur « logique » (cf. Figure 2.2) lorsqu'il n'est que pur discours, jeu de langage, réécriture (conformément à des règles).

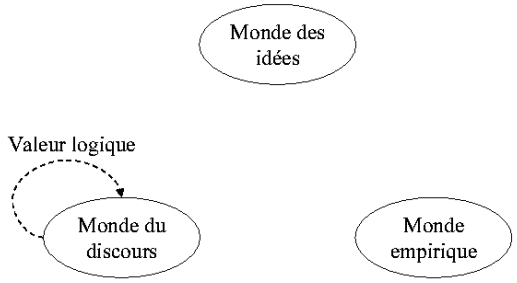


Figure 2.2 : Valeur logique du discours

En termes kantiens, il s'agit d'un cas particulier de raisonnement « *a priori* » : non seulement il est antérieur à toute observation, mais encore, dans ce cas précis, aucune observation ne sera nécessaire.

Par exemple, dans le cadre de référence de l'arithmétique des entiers, on peut récrire *1+1* en le remplaçant par *2* sans se soucier de savoir de quoi il s'agit. Des applications de cette règle peuvent même être contraires à l'expérience[24], ces applications se-

[24] Par exemple, dans le cas où l'on mélange un litre d'alcool et un litre d'eau, la proposition « un litre de liquide plus un litre de liquide font deux litres de liquide » est fausse.

ront alors réfutées mais pour autant les règles ne le seront pas. Cette règle de réécriture, purement formelle, opère sur un discours dé-corrélé de tout contenu.

b. L'obsession de l'interopérabilité

Un statut purement logique des modèles informatique signifierait que seul importe l'aspect formel, indépendamment du contenu (du rapport au monde). Dit autrement, il s'agirait de modèles qui puissent faire l'objet d'une normalisation.

Cette approche nous vaut, au passage, des expressions assez curieuses (pour ceux qui prendraient « ontologie » dans son sens traditionnel) comme « Standard Upper Ontology » (Groupe de travail IEEE) ou « Formal Ontology » [25].

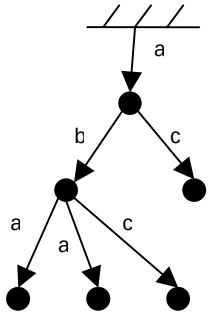
> The general perspective I have in mind is that of Formal Ontology, which can be included as the theory of formal distinctions between the elements of a domain, independently of their actual reality. [Guarino, 1997]

Quels peuvent être les vertus de la normalisation si ce n'est de permettre à différents systèmes informatiques de travailler ensemble ? Les tenants de cette approche normalisatrice le reconnaissent d'ailleurs parfois eux-mêmes : « "The Semantic Web is really data that is processable by machine" says Berners-Lee » [Frauenfelder01].

Cette révision à la baisse des objectifs (de l'intelligence artificielle vers l'interopérabilité) apparaît d'ailleurs en filigrane par l'inflation d'intérêt autour des formats que sont XML (format de sérialisation d'un arbre, cf. Figure 2.3) et RDF (format de sérialisation d'un graphe, cf. Figure 2.4).

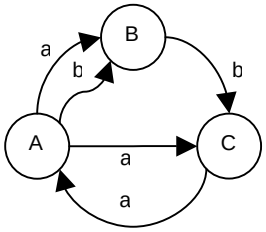
> Two important technologies for developing the Semantic Web are already in place: eXtensible Markup Language (XML) and the Resource Description Framework (RDF). [BernersLeeEtAl01]

[25] Bruno Bachimont faisait déjà remarquer [Bachimont01, dans son intervention orale] qu'« ontologie formelle » était un *oxymoron*, c'est à dire une expression alliant deux mots contradictoires.




```
<a>
   <b>
      <a/>
      <a/>
      <c/>
   </b>
   <c/>
</a>
```

Figure 2.3 : Extrait d'un exemple de sérialisation d'arbre en XML




```
<rdf:Description rdf:about = A >
   <a rdf.resource = B />
   <b rdf.resource = B />
   <a rdf.resource = C />
</rdf:Description>
<rdf:Description rdf:about = B > <b rdf.resource = C /> </rdf:Description>
<rdf:Description rdf:about = C > <a rdf.resource = A /> </rdf:Description>
```

Figure 2.4 : Extrait d'un exemple de sérialisation de graphe en RDF

Cependant, comme nous allons le voir maintenant, cette vision purement formelle des modèles informatiques ne doit nullement nous surprendre, étant donné l'histoire même de la discipline et des liens qu'elle entretient avec sa discipline mère : les mathématiques (modernes).

c. Courant philosophique

Au XIX° et au début du XX°, l'idéal de la Science est donné par la Mathématique[26] telle qu'elle serait si elle était libérée de l'influence du mathématicien : une mathématique « mécanique », une mathématique exécutable par une machine.

Hilbert, en refondant les mathématiques sur des axiomes et des inférences, suit ce programme, mais ses résultats vont aller à l'encontre des attentes des positivistes. En effet, on cite souvent la phrase d'Hilbert : « Il doit toujours être possible de substituer "table", "chaise" et "chope de bière" à "point", "droite" et "plan" dans un système

d'axiomes géométriques ». Si on y regarde de plus près, Hilbert, dans cette citation, coupe définitivement les mathématiques de la réalité. La vérité en mathématique n'est donc plus « ce qui est conforme au réel » mais ce qui est inféré d'une axiomatique. La vérité n'y est plus ontologique mais logique (au sens de « conventionnel »).

En fait, cette révolution a été préparée par les géométries non-Euclidiennes [Leonhardt99b]. Euclide a fondé la géométrie (principalement) sur cinq postulats. Mais, parce que le cinquième[27] semblait différent des autres, des légions de mathématiciens ont essayé de prouver qu'il était déductible des quatre autres. Aucun de ces mathématiciens n'avait le moindre doute concernant la vérité de ce postulat/théorème. N'était-ce pas vrai ontologiquement ? La réalité n'était-elle pas régie par la géométrie d'Euclide ? C'est alors que Lobachevsky (1829)[28] ; puis Riemann (1854) proposent chacun de remplacer le cinquième postulat par un postulat contradictoire[29]. En faisant cela, ils obtiennent deux géométries (dites non-euclidiennes) cohérentes quoique contraires à l'évidence.

On peut se demander si l'approche logique est auto-fondée ou si elle fondée ontologiquement. Pour répondre, nous devons étudier l'histoire du théorème d'incomplétude de Gödel (1931) [Dubucs96, Girard89]. Par ce théorème, Gödel prouve que pour toute axiomatique (aussi complexe ou plus complexe que l'arithmétique des entiers) soit le principe du tiers-exclu[30], soit celui de non-contradiction[31] est violé. Le plus surprenant est que Hilbert a passé le reste de sa vie à chercher une erreur dans la démonstration et que Gödel lui-même a toujours pensé que l'incomplétude était due aux systèmes complexes et pas à la logique « naturelle » (c'est-à-dire contenant les deux principes).

---

[26] L'usage du singulier a ici pour but de renforcer le projet d'unification de la discipline, projet qui va au contraire aboutir à un éclatement…

[27] Le cinquième postulat peut être exprimé par : "Par un point extérieur à une droite, passe une et une seule parallèle à cette droite".

[28] Ainsi que Bolyai en 1832.

[29] Respectivement : "Par un point extérieur à une droite, passent plusieurs parallèles à cette droite", et "Par un point extérieur à une droite, ne passe aucune parallèle à cette droite".

[30] "D'une proposition et de son contraire, au moins l'une est vraie" (les deux pouvant être vraies).

[31] "D'une proposition et de son contraire, au plus l'une est vraie" (les deux pouvant être fausses).

En fait, même s'ils ont défini les mathématiques comme conventionnelles, ils gardent tout deux une foi sans faille dans la nature ontologique de la logique. On doit noter, au contraire, que d'autres mathématiciens à la même époque plaident pour l'intuitionnisme (c'est-à-dire une logique sans le principe de non-contradiction).

En résumé, le statut des mathématiques a changé. Les mathématiques ne sont plus considérées comme conformes à la réalité. Elles sont un jeu de langage. Aujourd'hui, même les règles de la logique « naturelle » ne nous semblent plus « naturelles » et peuvent être changées afin d'obtenir de nouvelles logiques.

3. Valeur épistémologique

a. Définition

Le mot « épistémologique » est formé sur deux racines grecques :

- *episteme*, la connaissance,
- *logos*, le discours (raisonné).

Nous dirons qu'un discours a une valeur « épistémologique » (cf. Figure 2.5) lorsqu'il naît d'une idée antérieure à l'observation. En termes kantiens on parlerait de raisonnement « *a priori* ». La vérité de cette connaissance pourra être mise à l'épreuve en projetant le discours sur le réel.

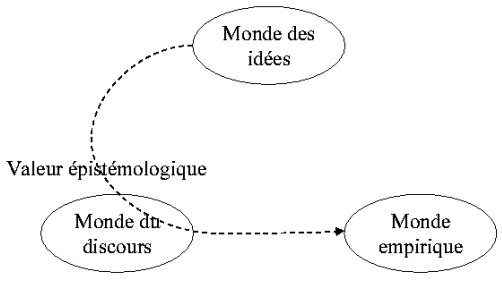


Figure 2.5 : Valeur épistémologique du discours

Par exemple, Newton, a eu l'idée de définir la masse (notion *a priori*) comme le rapport constant d'une accélération avec une force. Cette théorie, d'une étonnante effica-

cité, a fini par être mise à mal par l'expérience (lorsque les corps atteignent des vitesses proches de celle de la lumière).

b. Des machines qui donnent à penser

Si les modèles informatiques ont un statut épistémologique, cela signifie qu'ils n'ont de sens que par l'usage et l'interprétation qu'en feront leurs utilisateurs humains. Nous faisons ici écho aux thèses défendues par Jean Charlet, Bruno Bachimont et leur équipe :

> La tâche conditionne la construction de l'ontologie qui dès lors ne peut être portable et encore moins universelle[32]. Cela plaide de plus en faveur d'une vision non logique mais plutôt constructiviste de la connaissance [..]. Comme toute connaissance, les ontologies sont interprétées par un expert humain en fonction de l'idée qu'il a de la tâche attribuée au système. [CharletEtAl96]

C'est d'ailleurs ce qu'affirment les initiateurs du Web Sémantique après avoir soutenu les deux thèses contraires :

> The computer doesn't truly « understand » any of this information, but it can now manipulate the terms much more effectively in ways that are useful and meaningful to the human user. [BernersLeeEtAl01]

Alliant la théorie à la pratique, le projet du Web Sémantique dispose même d'un sous-projet appelé *Annotea* consistant à faire créer par des tiers des ontologies sous forme d'annotations. Et le projet de mentionner que ceci entraînera sans doute des « biais » (entre autres politiques[33]) ainsi qu'une notion de confiance, mais que c'est le prix à payer.

Dans le même esprit, John F. Sowa propose de baser les ontologies sur ce qu'il appelle « le principe de Socrate » :

---

[32] Pour s'en assurer, il suffit de consulter le catalogue pléthorique quoique incomplet des ontologies candidates à l'universalité qu'a dressé Fritz Lehmann [Lehmann94].

[33] Quelle valeur donner au fait que le pouvoir de Pékin définisse Taiwan comme une partie de la Chine ?

> Whenever two philosophers disagree, draw a distinction. Anyone may discover a conflict, Anyone may suggest a distinction, Machine recomputes [..], Repeat until everybody is happy. [Sowa01b]

Dit autrement, il considère les polémiques comme le moteur de la modélisation et la machine comme un support dynamique pour un discours en perpétuelle réécriture.

Notons que baser un système sur la polémique est un véritable défi informatique. En effet la plupart des « systèmes à base de connaissances » sont basés sur la logique classique. Cela signifie premièrement qu'ils ne peuvent exprimer deux points de vue contradictoires (principe de non-contradiction). Deuxièmement, cela signifie qu'ils considèrent un « monde fermé » dans lequel tout est connu (vrai ou faux) et ne laisse donc pas de place à l'indécision (principe du tiers exclu).

Nous ne pouvons ici qu'aller dans le sens du Web Sémantique (troisième mouture) qui a pris conscience du défi et l'accepte sereinement :

> Semantic Web researchers, in contrast, accept that paradoxes and unanswerable questions are a price that must be paid to achieve versatility. [BernersLeeEtAl01]

c. Courant philosophique

Au milieu du XX° siècle, Popper [Popper35], influencé par le Positivisme Logique, entreprend de fonder la Science de manière objective, c'est-à-dire, sur les expériences et la logique. Cependant, à la différence des positivistes logiques, il renonce à démontrer des propositions universelles affirmatives (lois) à partir de principes tirés de l'expérience. En effet, ce type de raisonnement, appelé « induction amplifiante »[34] est reconnu comme un raisonnement fallacieux depuis les débuts de la logique. Au contraire, il choisit la déduction (et plus précisément le *modus tollens*[35]), pour bâtir sa

[34] Exemple d'induction amplifiante : « Le cygne n°1 est blanc, le cygne n°2 est blanc, ... le cygne n°*N* est blanc, par conséquent tous les cygnes sont blancs ». A ne pas confondre avec l'induction simple : « Le cygne n°1 est blanc, le cygne n°2 est blanc, ... le cygne n°*N* est blanc, par conséquent tous les cygnes du numéro 1 au numéro *N* sont blancs ».

[35] « Si *p* est vrai alors *q* est vrai, or *q* est faux, par conséquent *p* est faux ».

méthode hypothético-déductive :

SI hypothèse testable est vraie

ALORS phénomène objectif apparaît

OR phénomène objectif n'apparaît pas

PAR CONSEQUENT hypothèse testable est fausse.

D'un point de vue systémique, la méthode hypothético-déductive peut être modélisée par un processus ayant comme entrée une « vérité hypothétique » et comme sortie une « vraie réfutation ».

Premièrement, on est en droit de se questionner sur la « vérité » de la méthode (et donc de ses réfutations) : ontologique, logique, épistémologique ? En fait ce point n'a jamais été abordé par Popper. Il semble qu'il ait pensé que le *modus tollens* était « naturel », « évident ». Avec le recul, on peut donc affirmer que sa méthode est enracinée dans des croyances ontologiques. Si l'on souhaite suivre la méthode popperienne tout en tenant compte des mathématiques modernes, nous devons remplacer sa « croyance » dans la logique « naturelle », par le choix, conventionnel, de cette logique parmi d'autres.

Deuxièmement, le processus a besoin de deux entrées : une « hypothèse testable » et un « phénomène objectif », mais rien ne nous dit dans la méthode comment les obtenir. Elle laisse une place pour quelque chose d'autre...

Concernant « l'hypothèse testable », Popper écrit que son invention par un être humain concerne la psychologie et non l'épistémologie (puisque cette invention n'interfère pas avec sa méthode logique). On lui sera reconnaissant de ne pas invoquer, contrairement à d'autres philosophes, d'obscurs processus formels tels que l'induction ou l'abduction, et de préférer garder une méthode purement logique tout en laissant de la place aux Sciences Humaines.

Concernant le « phénomène objectif », c'est un phénomène qui est expérimentalement reproductible indépendamment du chercheur. Combien de fois une expérience doit-elle être répétée ? Par combien de chercheurs ? Là encore, Popper n'a rien écrit à ce sujet. Par contre, Kühn (cf. [Kuhn62]) peut nous aider à trouver une réponse. Kühn donne de la Science une vue sociologique. Les chercheurs appartiendraient à des communautés basées sur des principes tacites. Ces principes (appelés « paradigmes ») établissent les méthodes à utiliser et les types de sujets de recherche solubles. Ainsi, des théories basées sur des paradigmes différents sont incommensurables (elles ne peuvent être comparées). D'une manière kühnienne, on pourrait affirmer qu'un « phénomène objectif » est objectif pour une communauté donnée si et seulement si il a été reconnu tel par cette communauté. Un « phénomène objectif » (et son pouvoir réfutant) n'est donc pas forcément valide pour une autre communauté. De même, il peut ne plus être valide après un changement de paradigme dans la communauté (ce que Kühn appelle une « révolution »). En définitive, « l'objectivité » des expériences et des réfutations dépend d'un processus social et en tant que tel est du ressort des Sciences Humaines.

4. Conclusion

Le lecteur aura compris que nous nous engageons clairement dans la troisième voie, celle que nous avons qualifiée « d'épistémologique ». Cela signifie que nous limitons la mission de l'ordinateur à ce qui est purement formel et laissons à la charge de l'être humain ce qui concerne la substance des choses. C'est parce que nous sommes convaincus que l'intelligence sera toujours du côté de l'humain et non de l'automate qu'il nous semble indispensable de mettre au cœur de la conception de nos systèmes : l'interactivité, les conflits d'interprétation ainsi que l'aspect dynamique et toujours incomplet de la connaissance.

Chapitre 3. Explication et compréhension en Sciences Humaines*

Deux dangers ne cessent de menacer le monde :
l'ordre et le désordre.
Paul Valéry

Dans le précédent chapitre, nous avons placé nos recherches dans une certaine conception de l'informatique : une informatique de l'interaction et de la communication ; une informatique dont les modèles sont construits de manière hypothétique, transitoire et polémique. Nous proposons ici, d'une part, d'étudier en quoi l'utilisation d'une informatique de ce type peut être judicieuse en Sciences Humaines (en particulier en archéologie). D'autre part, nous devons voir si les méthodes en Sciences Humaines ne pourraient pas nous orienter vers un type de modèles nous permettant d'atteindre les objectifs fixés.

1. Le cas de la formalisation en archéologie

Plusieurs dizaines d'années après les grands projets de bases factuelles, de systèmes experts et autres « banques de données », le chercheur en archéologie a toujours, comme principaux outils, les publications savantes. L'un des grands instigateurs de ces projets, Jean-Claude Gardin, reconnaît lui-même le peu d'intérêt que rencontrèrent ses banques de données archéologiques et interprète ces échecs comme étant dus à la difficulté de distinguer en archéologie les « faits », des conclusions ou des interprétations [Gardin86]. Pour aller plus loin, on pourrait se demander si le paradigme computationnel, qui a pourtant eu le succès qu'on lui connaît dans les entreprises, ne serait pas, tel quel, inapplicable à un certain nombre de domaines. La Médecine, le Droit, les Sciences de l'Homme et de la Société, sont autant de domaines où la recherche en informatique a

* Ce chapitre a fait l'objet d'un article de recherche à paraître dans les Suppléments au Bulletin de Correspondance Hellénique [Benel03].

dû s'adapter au seul support de la connaissance reconnu valide par ses praticiens : le document.

Sans même requérir l'argument épistémologique de l'inadéquation de la formalisation aux Sciences de l'Homme, il est impossible, comme le concède Jean-Claude Gardin[36], de parler de bases de données archéologiques sans s'interroger sur la reconnaissance de ces données par l'ensemble de la communauté des chercheurs en archéologie. Aussi, René Ginouvès et Anne-Marie Guimier-Sorbets [GinouvesEtGuimierSorbets78] écrivaient que, plutôt que d'espérer en un hypothétique consensus assurant l'objectivité des données, mieux valait viser l'intersubjectivité. De même, pour Patrick Desfarges et Bruno Helly [DesfargesEtHelly91] : « Les objets n'ont pas d'attributs par eux-mêmes mais par leurs sources », principe mis en pratique avec le système FRANTIQ, dans lequel sont enregistrés des « discours » d'auteurs sur des artefacts et non des données impersonnelles.

En effet, à la différence des bases de données, le document daté, authentifié et soumis à un comité éditorial est en adéquation totale avec une science moderne de type constructiviste, c'est à dire où la scientificité repose sur la « réfutabilité » d'hypothèses par l'expérience [Popper35] et/ou par les pairs [Kuhn62].

Si les documents semblent adaptés à la pratique scientifique, il en est autrement des systèmes documentaires courants. Conçus pour des bibliothèques grand-public, la plupart reposent sur une indexation effectuée par des tiers selon une structure statique (ontologie, thésaurus, liste d'autorité). Or, structurer le corpus de documents – et par là le domaine – à la place du chercheur, c'est nier son expertise. En effet, c'est la mission même du chercheur de trouver une structure à son objet d'étude et de tester la validité de cette structure. Par conséquent, un système documentaire adapté aux chercheurs ne devrait pas être basé sur des « méta-données » fixes, mais être le support d'une activité exploratoire de structuration.

---

[36] « Par quels mécanismes obtiendra-t-on que l'accord initial [..] engageant une population de chercheurs limitée dans l'espace et le temps, s'étende ensuite de façon quasi-statutaire [..] ? » [Gardin84]

2. L'indispensable compréhension

Afin de dresser un aperçu de l'épistémologie des Sciences Humaines, il nous est nécessaire de préciser tout d'abord ce que l'on entend par « Sciences Humaines ». L'œuvre de Wilhelm Dilthey (fin XIX° s. – début XX° s.) est en ce sens très éclairante [Ricoeur86]. Elle propose en effet un critère de démarcation entre les Sciences de la Nature et les Sciences Humaines (qu'il appelle Sciences de l'Esprit). Pour lui, ce critère repose sur la dialectique entre explication et compréhension. L'explication serait le mode de construction des Sciences de la Nature, tandis que la compréhension serait celle des Sciences de l'Esprit.

L'explication concerne le « comment », la compréhension le « pourquoi ». Alors que le premier est l'étude d'un objet indépendamment de l'observateur (objectivité), la seconde est l'étude d'un sujet (une personne) par un autre sujet (subjectivité). L'Explication nécessite une mise à distance (distanciation). A l'inverse la compréhension (« prendre avec ») requiert une appropriation (actualisation). La première manipule de la signification, propriété de signes dé-contextualisés[37], à la différence de la seconde qui manipule du sens, propriété de signes en contexte et en situation (cf. François Rastier sur la dialectique Sens/Signification[38] ; [Ducrot72] sur la dialectique Contexte/Situation).

On entend par « contexte d'un signe », étymologiquement parlant, ce qui est « tissé avec », c'est-à-dire l'ensemble des signes qui l'entourent. Un texte est ainsi, pour reprendre son étymologie, un « tissu ». La situation, quant à elle, étend la notion de contexte à un « tissu » également non linguistique (temps, lieu, personne...). Pour un texte donné, peuvent être considérées à la fois les situations d'écriture et de lecture. La

[37] Par exemple, dans le cadre de référence de la Physique classique, « masse » et « vitesse » possèdent une signification fixe quels que soient leurs contextes d'apparition.

[38] « la signification est une propriété assignée aux signes, le sens une propriété des textes. [..] la signification résulte en effet d'un processus de décontextualisation [..]. En revanche le sens suppose une contextualisation maximale aussi bien par la langue (le contexte, c'est tout le texte) que par la situation (qui se définit par une histoire et une culture, au-delà du *hic et nunc* de la pragmatique). » [Rastier98, p.7-8]

première, par exemple, comprend la culture et l'histoire à la fois de l'auteur et de la société auxquels le texte appartenait.

Afin d'éviter de plonger les Sciences Humaines dans une subjectivité débridée, Dilthey choisit de placer la compréhension dans la tradition méthodologique de l'Herméneutique, c'est à dire la science de l'interprétation. En effet, cette science destinée au départ à l'étude des textes sacrés, avait posé la question des limites de l'interprétation et proposé un certain nombre de réponses : étudier la situation dans laquelle un texte a été écrit, comparer un texte avec des textes contemporains ou antérieurs, avec d'autres versions du même texte, avec des commentaires postérieurs, multiplier les interprétations et engager le débat, ou en dernier recours s'en remettre à la tradition d'interprétation, à l'autorité, ou au « préjugé » pour se forger une opinion provisoire. En résumé, avec Dilthey, apparaissait une épistémologie assez claire des Sciences de l'Esprit, épistémologie guidée par une herméneutique de la compréhension, totalement disjointe du processus explicatif (cf. Figure 3.1).

Ce modèle va être remis en cause avec la naissance du structuralisme en linguistique au début du XX° siècle et surtout son extension à toutes les spécialités des Sciences Humaines dans les années 1960 et 1970. En effet, ce mouvement à tendance clairement explicative va s'avérer d'une fécondité remarquable (en anthropologie, psychanalyse, critique littéraire, étude de la grammaire, etc.) réfutant l'omission de la dimension explicative dans l'épistémologie des Sciences Humaines. A l'inverse, la tentation de réduire les Sciences Humaines à la seule explication s'est avérée à l'expérience comme désastreuse.

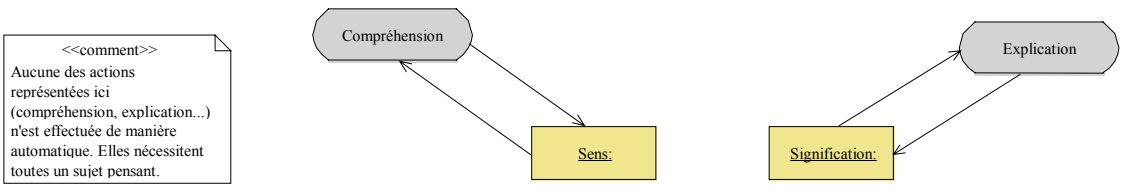


Figure 3.1 : L'interprétation d'après Dilthey comme compréhension pure – disjointe de l'explication (diagramme d'activité UML)

La thèse centrale de Paul Ricœur dans son deuxième essai d'herméneutique [Ricoeur86] est de rassembler sous la méthode herméneutique les dimensions de la compréhension et de l'explication ; de considérer l'interprétation non plus comme la seule compréhension mais comme la dialectique même de la compréhension et de l'explication. En effet, la distanciation nécessite l'appartenance préalable à des sujets, et à l'inverse il n'existe pas de compréhension de soi sans médiation par des signes, symboles ou textes[39].

En résumé, pour Paul Ricœur, le texte est le trait d'union entre les processus de compréhension et d'explication ; l'ensemble de ces deux processus étant régi par les méthodes de l'herméneutique[40] (cf. Figure 3.2).

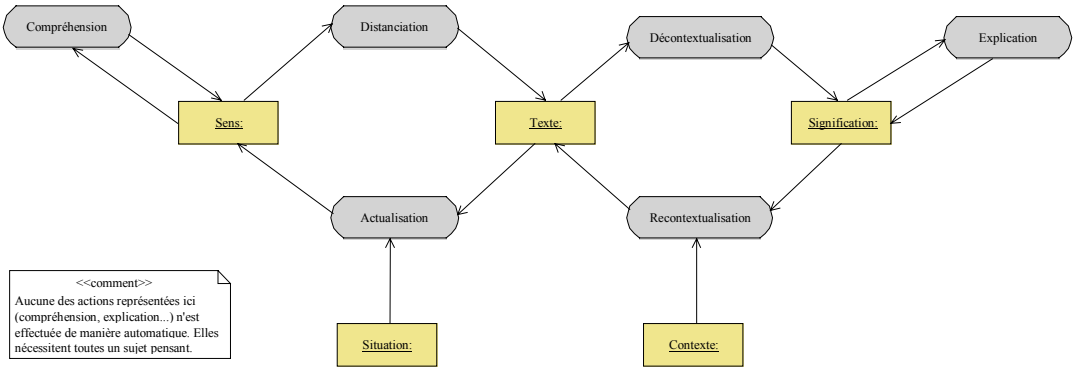


Figure 3.2 : L'interprétation d'après Paul Ricœur comme compréhension et explication mises en relation par le texte (diagramme d'activité UML)

[39] « Ce combat sur deux fronts contre une réduction de la compréhension à l'intropathie et une réduction de l'explication à une combinatoire abstraite m'amène à définir l'interprétation par cette dialectique même de la compréhension et de l'explication au niveau du "sens" immanent au texte. » [Ricoeur86, p.37-38]

[40] « quelle peut être la tâche première de l'herméneutique ? Elle est selon moi, de chercher dans le texte lui-même, d'une part la dynamique interne qui préside à la structuration de l'œuvre, d'autre part la puissance de l'œuvre de se projeter hors d'elle-même et d'engendrer un monde qui serait véritablement la "chose" du texte. Dynamique interne et projection externe constituent ce que j'appelle le travail du texte. C'est la tâche de l'herméneutique de reconstruire ce double travail du texte. » [Ricoeur86, p.36]

3. Un cadre interprétatif pour le calcul

Reste à définir la place du calcul (c'est-à-dire de l'informatique) dans notre modèle. Le calcul apparaît clairement comme faisant part du processus d'explication. Cependant, des précisions s'imposent. En effet, alors que l'explication, comme nous l'avons vu, porte sur des significations, le calcul porte sur des symboles (au sens mathématique) c'est-à-dire des objets de calcul sans aucune signification. Ioannis Kanellos parle de « forme décorrélée de tout contenu » [KanellosEtAl00] ; Bruno Bachimont de signe autothétique (étymologiquement, « qui porte soi-même »), c'est-à-dire « qui ne représente rien sinon sa propre effectivité » [Bachimont99b]. Là où les choses se compliquent, c'est que les usagers des systèmes informatiques attribuent une signification aux symboles (et même plus puisqu'ils sont, selon l'expression de François Rastier, « condamnés au sens[41] »).

Bruno Bachimont considère que le symbole autothétique se voit surdéterminé jusqu'à devenir orthothétique (étymologiquement, « qui porte exactement »), c'est-à-dire un enregistrement de la parole telle que l'écriture alphabétique[42].

Ioannis Kanellos en vient à définir deux processus humains à savoir la désémiotisation et la re-sémiotisation qui permettent de passer respectivement de la signification au symbole et du symbole à la signification [KanellosEtAl00].

Prenons un exemple. Considérons un chercheur travaillant sur le dieu Hermès. En cherchant dans un corpus de textes le nombre d'occurrences de l'expression régulière « HERM* », il dé-sémiotise une signification en une suite de symboles informatiques.

---

[41] « les phrases réputées absurdes, voire asémantiques peuvent toujours être interprétées […] on peut lire Finnegan's Wake même dans les passages où aucun des mots ne figurent au dictionnaire […] C'est là une allégorie du péché originel, ou du moins de la condition humaine : nous sommes condamnés au sens. » [Rastier91, p.212-213]

[42] D'où son projet de considérer l'intelligence artificielle non plus comme une modélisation au sens fort mais comme une écriture dynamique à interpréter [Bachimont99a]. Notons qu'il s'agit de l'approche inverse de celle de Jean-Claude Gardin qui, pour éviter les écueils supposés de l'interprétation, préconisait d'appliquer le formalisme des systèmes experts aux publications savantes en Sciences Humaines [Gardin86].

Le système calcule une liste de réponse. Le chercheur effectuera ensuite une re-sémiotisation, en interprétant par exemple la faible fréquence des occurrences comme une certaine discrétion du dieu dans le panthéon. Notons que dé-sémiotisation et re-sémiotisation ne sont jamais neutres. Par exemple le fait de compter toutes les occurrences de « HERM* » relève du choix du chercheur de considérer l'emploi métonymique de « hermès » (buste quelconque surmontant un support quadrangulaire) comme également représentatif de l'influence du dieu.

Ainsi, le processus explicatif se décompose en un processus humain de dé-sémiotisation permettant de transformer la signification en un symbole sémiotiquement neutre, puis d'un processus automatique de calcul, et enfin d'un processus humain de re-sémiotisation afin de donner une signification au résultat du calcul (cf. Figure 3.3).

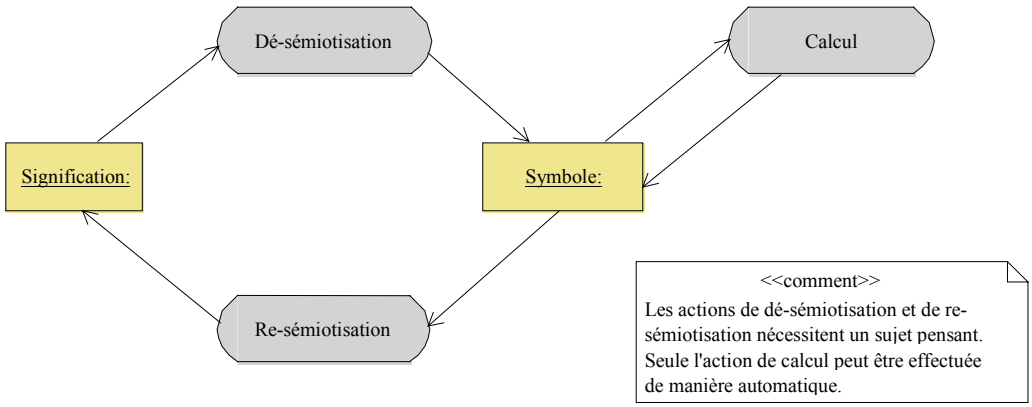


Figure 3.3 : L'explication par le calcul d'après François Rastier : le passage obligé de la signification au symbole – et inversement (diagramme d'activité UML)

En résumé, les processus de construction de sens sur un support informatique par un humaniste nécessitent des symboles qui soient à la fois sémiotisés, mis en contexte et mis en situation. Autrement dit, l'instrumentation du travail des chercheurs en Sciences Humaines passent par la constitution dynamique d'un *corpus* : une sorte de « dossier numérique ». Comme l'ont analysé Jean Charlet et son équipe pour le domaine médical, le dossier numérique permet par sa forme documentaire, son aspect contextuel, et la pluralité des auteurs et des lecteurs qui le manipulent de rendre compte d'un domaine où les objets (voire les sujets) d'étude sont complexes et où les modèles évoluent [Charle-

tEtAl99]. Quant à nous, nous focaliserons notre attention sur deux aspects de ce dossier numérique :

- *l'intersubjectivité* : le dossier étant le « lieu » des conflits d'interprétation entre experts,
- *l'intertextualité* : la mise en contexte de fragments documentaires permettant de faire sens (en exprimant entre autres les accords et les conflits d'expertise).

Chapitre 4. Instrumentation de l'intertextualité et de l'intersubjectivité

Dans le précédent chapitre, nous avons vu comment les méthodes en Sciences Humaines recommandaient la constitution de corpus comme « lieux » d'intersubjectivité et d'intertextualité. Nous allons maintenant nous interroger sur l'instrumentation de ces deux notions à l'aide de systèmes informatiques. Tout d'abord, nous verrons que ces notions sont au cœur des usages (dans les bibliothèques traditionnelles). Ensuite, nous étudierons comment elles ont été modélisées dans le cadre de systèmes hypermédia.

1. Au cœur des usages

Instrumenter un travail n'est jamais une opération neutre. Comme le rappelle Bruno Bachimont [Bachimont99a] n'importe quel outil (du marteau au système informatique) détermine par sa structure des usages possibles (ce qui n'empêche pas des usages déviants). La question de l'adéquation de l'outil au travail à effectuer est donc déjà sensible au niveau d'une entreprise. Elle devient flagrante dans le cas de la recherche publique, pour laquelle, il serait tout à fait déplacé de la part d'une instance hiérarchique d'imposer l'utilisation d'un outil.

Aussi, certains pourraient être surpris par notre choix d'élaborer un modèle du travail à partir de textes méthodologiques (censés le guider) plutôt qu'à partir d'un travail « d'anthropologie » (immersion dans une communauté, observation et entretiens). Cependant, il faut constater que les études « anthropologiques » menées par d'autres équipes corroborent tout à fait l'utilisation de l'intertextualité et de l'intersubjectivité.

Ainsi, Kenton O'Hara et son équipe [OHaraEtAl98] ont étudié les activités documentaires effectuées par des doctorants durant une journée de travail. Ils décrivent le travail des usagers des bibliothèques universitaires comme allant bien au-delà de la traditionnelle recherche d'information bibliographique. En effet, devraient être considérées comme formant un tout :

- la lecture,
- l'annotation,
- la rédaction de fiches de lecture,
- la reproduction par photocopie,
- la consultation d'une bibliographie,
- la relecture,
- l'écriture de nouveaux documents (articles, mémoires, bibliographies…).

De même, Jacques Virbel et son équipe, à la suite d'une collaboration avec les grands lecteurs de la Bibliothèque Nationale de France, ont mis en évidence huit verbes d'action liés à la lecture active de documents [MazhoudEtAl95, Veron97] :

- Hiérarchiser (attribuer un niveau d'importance),
- Architecturer (typer des fragments),
- Contextualiser (borner le sens),
- Reformuler,
- Commenter (critique, idée associée),
- Documenter (ajouter un fragment pour en comprendre une autre),
- Corréler (relier deux fragments),
- Programmer (projeter une action : à traduire, à relire, à analyser…).

Enfin, dans le centre de documentation d'une entreprise, Andreas Paepcke [Paepcke96] a interviewé des ingénieurs sur leurs besoins et pratiques informationnelles. Il en ressort que même si la recherche d'information est centrale, celle-ci est corrélée avec trois actions nécessitant la communication entre des êtres humains, à savoir :

- la découverte,
- la gestion
- et le partage de ces informations.

Après de tels constats, on pourrait s'étonner du fait que très peu de systèmes de bibliothèques numériques permettent les interactions sociales [Tochtermann96] et l'enrichissement par le lecteur [GohEtLeggett00, RoscheisenEtAl95, NanardEtNanard00]. Cependant, comme nous allons le voir, ces thèmes ont été en partie traités par une autre communauté : celle des Hypermédia.

2. Panorama des systèmes hypermédia

Sans prétendre à l'exhaustivité, nous aimerions donner une vue aussi représentative que possible de différents types de modèles hypermédia. La documentation des modèles sous-jacents étant souvent insuffisante, nous avons dû effectuer une « rétro-conception » à partir des bribes de modèles donnés par les auteurs, à partir d'exemples d'utilisation, ainsi que de copies d'écran.

a. Xanadu®

Xanadu® est à juste titre appelé « the original hypertext project ». En effet, il s'agit du projet que poursuit, depuis les années 1960, Ted Nelson (alors licencié en philosophie et maître en Sociologie), inventeur des termes « hypertexte » et « hypermédia ». L'idée originale consiste à lire deux textes en parallèle en exhibant les relations d'analogie (indiquées par le lecteur) ainsi que les relations de citation (indiquées par l'auteur). Notons que ces inclusions par référence (appelées « transclusions »), par opposition à des inclusions par copie, permettent la citation tout en évitant le plagiat [Nelson99].

Destiné au départ au texte, le modèle a été récemment étendu aux images. Si on tentait de généraliser l'approche, on pourrait dire qu'un *document* est un ensemble ordonné de *fragments* et de *liens* réutilisables (cf. Figure 4.1).

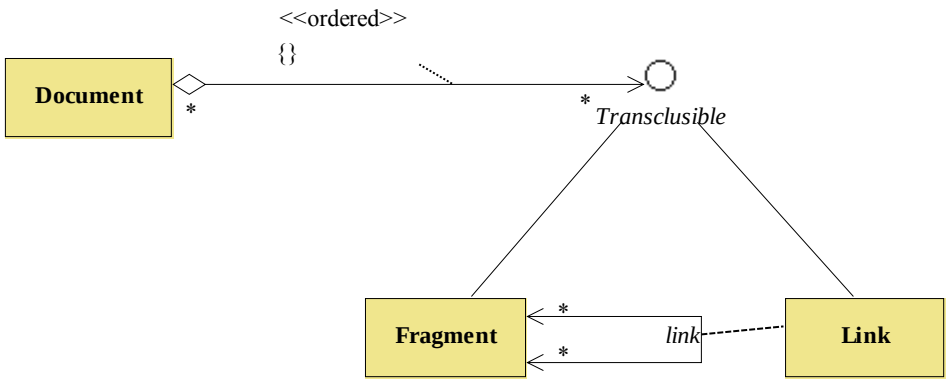


Figure 4.1 : Rétro-conception de *Xanadu*® (diagramme de classe UML)

On remarquera que le modèle ne propose des relations de composition que sur un seul niveau, ce qui aura pour conséquence des réseaux intertextuels soit verbeux soit incomplets.

Notons que l'expression d'un point de vue subjectif est tout à fait possible par la création de nouveaux documents. Cependant, l'instrumentation de l'intersubjectivité se limite, lors de l'affichage de deux textes, au soulignement de fragments corrélés : fonctionnalité fort intéressante mais sûrement pas suffisante.

b. World Wide Web

Présenter aujourd'hui le *World Wide Web* comme un hypertexte est sans doute un truisme, tant il est responsable de la démocratisation de la notion. Pourtant, il est intéressant de s'y arrêter un peu. Tout commence par un projet interne au CERN –projet mené par Tim Berners-Lee [BernersLee89]– ayant pour but d'unifier l'accès aux différentes bases de données du centre ainsi qu'aux rapports et publications scientifiques.

Ce projet est en fait basé sur un modèle de document hypermédia extrêmement simplifié (cf. Figure 4.2), limité à :

- des liens unidirectionnels (donc sans gestion de la cohérence par un serveur central),
- la transclusion d'images entières (mais pas de fragments),
- la transclusion de textes entiers (par l'intermédiaire des « cadres »).

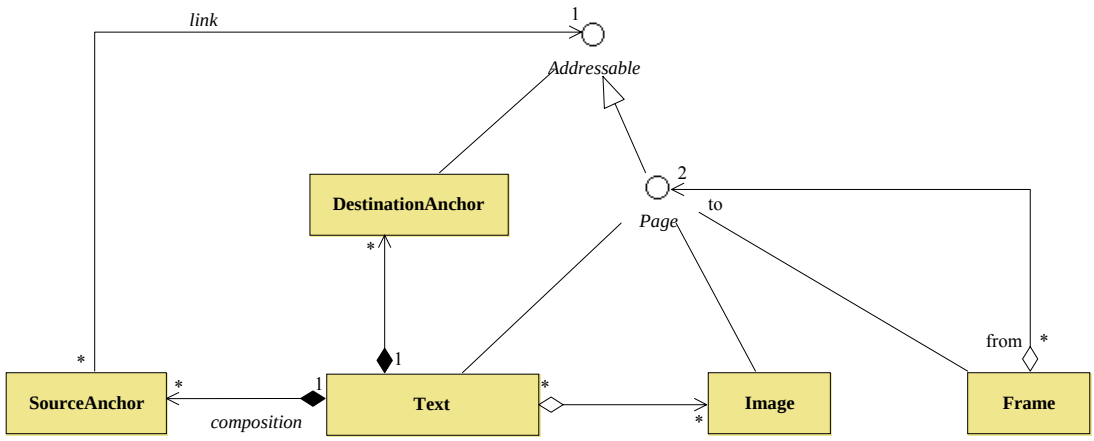


Figure 4.2 : Rétro-conception du *World Wide Web* (diagramme de classe UML)

Notons que c'est justement cette simplicité qui a permis au « Filet » (« Mesh »), interne au CERN, de passer à l'échelle d'une « Toile Mondiale ».

Si le succès commercial et l'engouement immodéré des « internautes » pour la Toile est indubitable (au moins à l'heure où nous écrivons), il en est autrement de son usage savant. Dans un manifeste « en ligne », Ted Nelson, rêvant toujours de « machines [pour les] littéraires », décrit la Toile comme une atteinte à l'intégrité de son invention :

> The Web isn't hypertext [..]. The Xanadu project has endeavored to implement a pure structure of links and facilitated re-use of content in any amounts and ways, allowing authors to concentrate on what mattered. What we have instead is the vacuous victory of typesetters over authors, and the most trivial form of hypertext that could have been imagined. [..] Instead, today's nightmarish new world is controlled by "webmasters", tekkies unlikely to understand the niceties of text issues [..]. [Nelson02]

Dans une perspective un peu plus constructive, Michael Bieber et ses collègues [BieberEtAl97] écrivent que la Toile est aux hypermédia ce que « l'assembleur » est aux langages de programmation, c'est-à-dire seulement la « seconde génération ». Dit autrement, la Toile n'offre que des fonctionnalités de « bas niveau », difficilement utilisables pour des projets ambitieux. Par contre, l'ensemble de ces fonctionnalités est suffisant pour en construire de plus haut niveau (cf. des projets comme HyperNietzsche

[IorioEtTurner99, Iorio00] ou NDWeb[43]). Cette idée n'était sans doute pas absente du projet originel de Tim Berners-Lee, puisque celui-ci, dès le départ, avait conçu les pages Web comme pouvant être calculées dynamiquement.

Concernant l'intertextualité, on pourra regretter à l'usage que le modèle ne permette de représenter la transclusion de plusieurs textes que par un arbre binaire dont les nœuds seraient des « cadres » et les feuilles seraient les pages. Ces « cadres » s'avèrent finalement difficilement utilisables (sans même parler de leur représentation graphique dans les navigateurs qui en a fait le cauchemar des créateurs de sites Web).

Pour ce qui est de l'intersubjectivité, on remarquera que si chacun peut créer une ressource subjective avec transclusion ou référence à des ressources existantes, il n'y aura aucun moyen (avec les outils standards) de remonter de la source à son commentaire : ce qui limite de beaucoup les possibilités d'interactions sociales.

c. Hyper-G

Hyper-G fut conçu à l'université technologique de Graz (Autriche) par une équipe dirigée par Hermann Maurer. Le système fit l'objet de plusieurs « doctorats en sciences techniques », en particulier celui de Frank Kappe, en 1991, portant sur le serveur [Kappe91] et celui de Keith Andrews, en 1996, sur le client graphique (appelé « Harmony ») [Andrews96]. Aujourd'hui, le système est devenu un produit commercial de gestion de connaissance appelé « HyperWave »[44] [Maurer96].

Hyper-G se distingue de la Toile en gérant (cf. Figure 4.3) des *liens* bidirectionnels dont la cohérence est maintenue (absence « d'erreur 404 »), ainsi que par des primitives hypermédia de plus haut niveau appelées « *collection* ». Une collection permet de contenir plusieurs *documents* (*documents atomiques* ou autres collections). Un même document pouvant appartenir à plusieurs collections, il est possible de représenter une structure non hiérarchique. Précisons également qu'il existe deux types de collections, à savoir la *séquence* et la *grappe*, suivant que la collection est ordonnée ou pas. Notons

[43] http://www.ndweb.org/recit/temple

[44] http://www.hyperwave.com

que le document atomique porte bien son nom puisqu'il est impossible de réaliser une transclusion d'une partie de cet objet. Ceci est d'autant plus regrettable que la définition de liens, par contre, utilise des parties de documents appelées « *ancres* ».

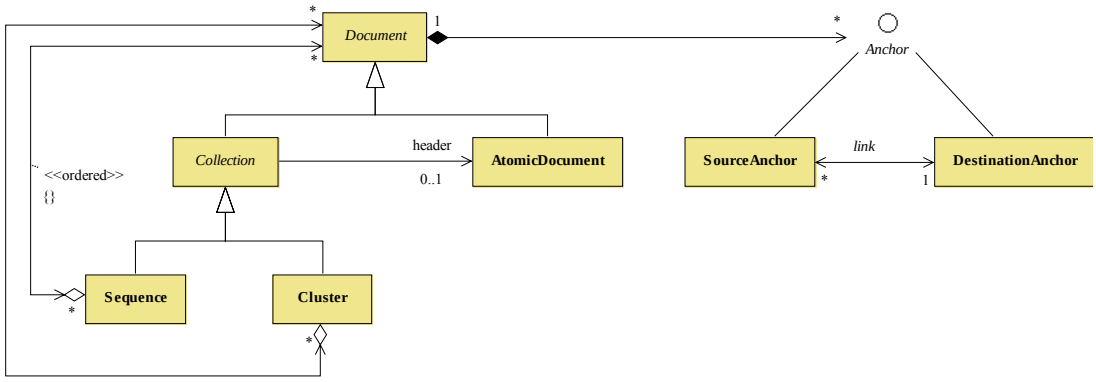


Figure 4.3 : Rétro-conception d'*Hyper-G* (diagramme de classe UML)

d. ATLAS.ti®

Le système *ATLAS.ti*® est issu d'un projet de recherche du département de psychologie de l'Université Technique de Berlin (1989-1992). Depuis 1993, il est commercialisé par son auteur, Thomas Muhr, en tant qu'atelier d'analyse qualitative de documents [Muhr97].

Passé pratiquement inaperçu dans notre discipline[45], le modèle d'*ATLAS.ti*® n'en demeure pas moins intéressant (cf. Figure 4.4). Son élément central est la citation, fragment défini par le lecteur sur un *document primaire*. Chaque citation peut être reliée à d'autres par l'usage *d'hyperliens*, et décrite par des *codes*. Ces codes, communs à plusieurs citations, peuvent être reliés à d'autres par des liens typés (cause, équivalence, généralisation, …). Un autre objet, le *mémo*, est un petit texte permettant de commenter un code, une citation ou un document primaire. Mémos, documents primaires, et codes peuvent être regroupées dans plusieurs *familles*. Enfin, le *supercode* se distingue du code par une définition en intension (en fonction d'autres codes, supercodes ou familles) des citations qu'il décrit.

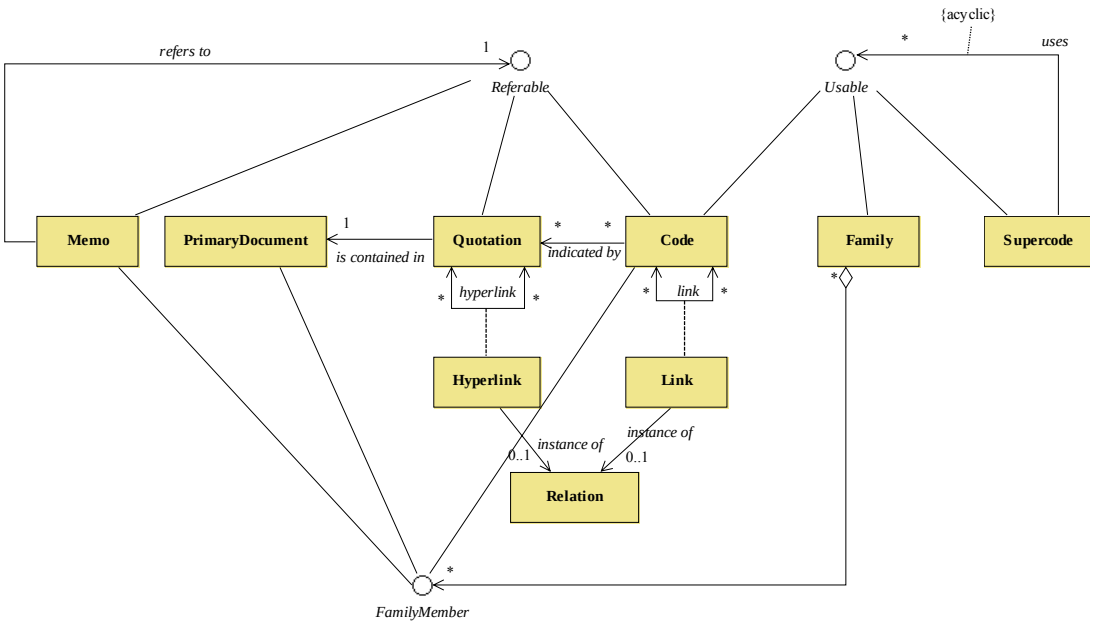


Figure 4.4 : Rétro-conception d'*ATLAS.ti*® (diagramme de classe UML)

Le modèle intertextuel présenté ici est le fruit d'une excellente connaissance de la pratique des chercheurs en Sciences Humaines. Il souffre cependant de quelques défauts de conception, notamment au niveau des généralisations. Ceci rend impossible un certain nombre d'actions qui sembleraient assez naturelles (coder un document primaire, commenter une famille ou un supercode, regrouper en une famille des citations ou des supercodes…) et complexifie inutilement le modèle.

Même s'il est possible de modéliser plusieurs points de vue en autant de « familles », on ne peut réellement parler ici d'intersubjectivité. En effet, la nature monoposte de l'application n'est guère propice au débat.

e. PASTEL

L'application *PASTEL* fut développée dans le cadre de la thèse de Ludovic Tanguy [Tanguy97a], thèse encadrée par Ioannis Kanellos et soutenue en 1997 à l'ENST de Bretagne.

---

[45] Merci à Ioannis Kanellos de nous l'avoir signalé.

Ce « Programme d'Aide à l'Analyse Sémantique de TExtes, même Littéraires » s'appuie sur un modèle informatique [Prie95, TanguyEtThlivitis96] de la sémantique interprétative de François Rastier (cf. Figure 4.5). L'interprète humain peut indiquer dans le texte à analyser un certain nombre de *sémèmes*. Les *taxèmes*, permettent de re-grouper en classes plusieurs de ces sémèmes, chaque sémème ne pouvant appartenir qu'à un seul taxème. Les *spécèmes*, pour leur part, permettent d'opposer deux à deux les sémèmes. Enfin, les *sèmes* permettent de qualifier spécèmes et taxèmes, ainsi que d'entrer en relation d'*afférence* avec les sémèmes.

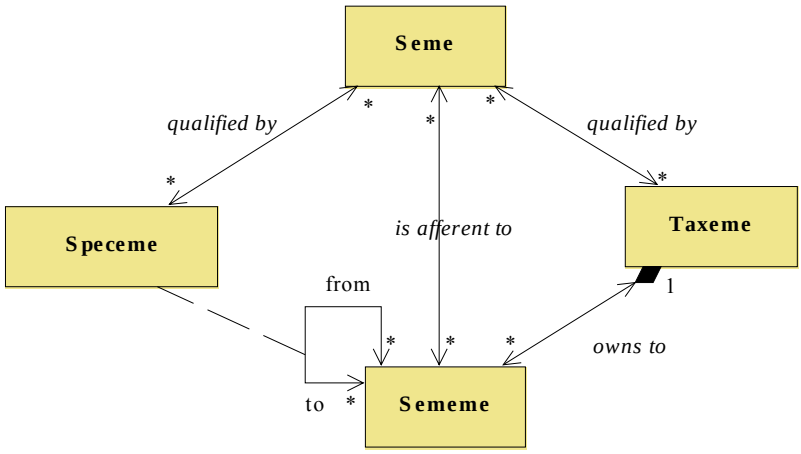


Figure 4.5: Rétro-conception de *PASTEL* (diagramme de classe UML)

De par son origine, ce modèle intertextuel est particulièrement approprié à l'analyse d'*un* texte. Par contre, tel quel, il est difficilement utilisable pour modéliser les différentes structures d'un corpus (en considérant que toutes sont « sémantiques »). En effet, le choix d'une sémantique purement différentielle, nous empêche de définir des relations générales (entre sèmes). Par exemple, au lieu d'indiquer une fois pour toute que la *Chronique des fouilles* se situe dans le *Bulletin de Correspondance Hellénique*, on serait obligé de préciser pour chaque passage du corpus[46] qu'il se situe dans les deux. Nous risquons donc d'obtenir, pour un corpus conséquent, un réseau aussi verbeux que celui de *Xanadu*®.

[46] Plusieurs dizaines de milliers.

f. Strates-IA

Le modèle *Strates-IA* est issu de la thèse de Yannick Prié [Prie99], thèse préparée au LISI sous la direction d'Alain Mille et de Jean-Marie Pinon, et soutenue en 1999.

Ce modèle (cf. Figure 4.6) est dédié à l'exploitation contextuelle d'*unités audiovisuelles* (fragments de *flux audiovisuels*). Le vocabulaire de l'annotation est défini par des *éléments d'annotation abstraits* reliés par des *relations conceptuelles.* Chaque relation conceptuelle peut éventuellement être déclarée comme instance d'un *type de relation*. L'instance d'un élément d'annotation abstrait dans une unité audiovisuelle est appelée « *élément d'annotation* ». Entre deux éléments d'annotation, il est possible de définir un lien (orienté) nommé « *relation élémentaire* ».

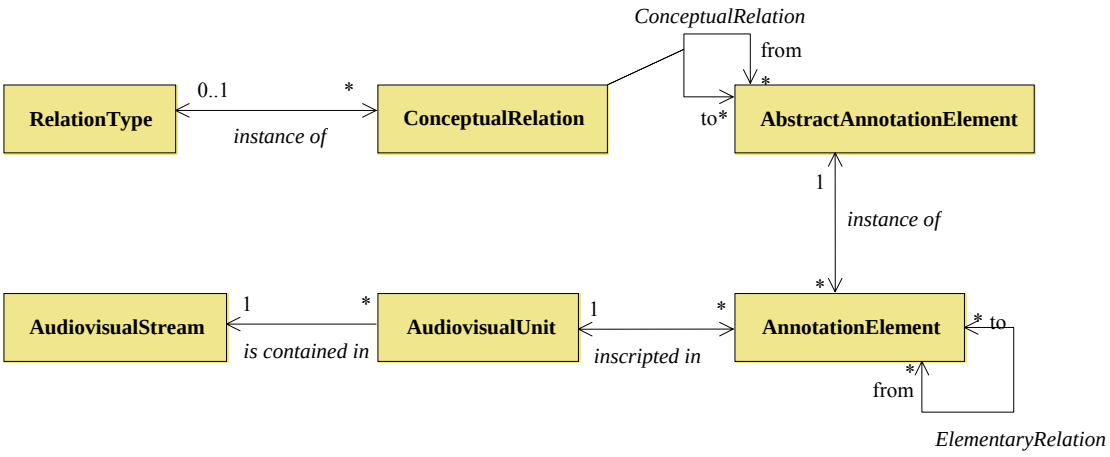


Figure 4.6: Rétro-conception de *Strates-IA* (diagramme de classe UML)

Si le modèle *Strates-IA* constitue un réseau intertextuel très réussi, l'expression de l'intersubjectivité sera par contre plus problématique. Comme l'équipe de *Strates-IA* et la nôtre l'avons expliqué dans un article commun [BenelEtAl01a], le rapport de l'individu à la connaissance diffère totalement dans nos domaines d'application respectifs. En effet, *Strates-IA* étant destiné à des documentalistes, il est naturel de séparer les autorités (éléments d'annotation abstrait) des indexations elles-mêmes (éléments

d'annotation). Dans notre perspective par contre, l'expert est amené à modifier les deux de conserve. En d'autres mots, il n'est pas question qu'il suive *le* modèle du domaine mais plutôt qu'il propose *son* modèle.

Notons que l'extension de *Strates-IA* réalisée par Elöd Egyed-Zsigmond dans le cadre de sa thèse, prend en compte les notions (définies informellement dans la thèse de Yannick Prié) de *dimension d'analyse* et de *schéma de description.* Ces deux notions peuvent être utilisées pour définir différents canons d'indexation. Si ceci est un premier pas vers la notion de point de vue, il s'agit tout de même de points de vue normés par une communauté et devant être appliqués par les individus. Il ne s'agit donc pas à proprement parler d'intersubjectivité.

g. TheBrain®

Le concept de *TheBrain*® [TheBrain01] fut inventé en 1994 par Harlan Hugh, un autodidacte en informatique alors âgé de 19 ans. La société[47] dont il est aujourd'hui président commercialise le produit en tant que « plateforme de connaissance ». Ce produit est protégé par une vingtaine de brevets (américains, autrichiens, italiens et européens).

Le modèle de *TheBrain*® est à la fois simple et général. Il est basé sur trois notions : celle de *contenu*, celle de *pensée* et celle de *relation* (cf. Figure 4.7). A chaque pensée, il est possible d'associer un contenu. Les pensées sont associées entre-elles par des relations de *paternité/filiation* ou de *saut*.

[47] http://www.thebrain.com

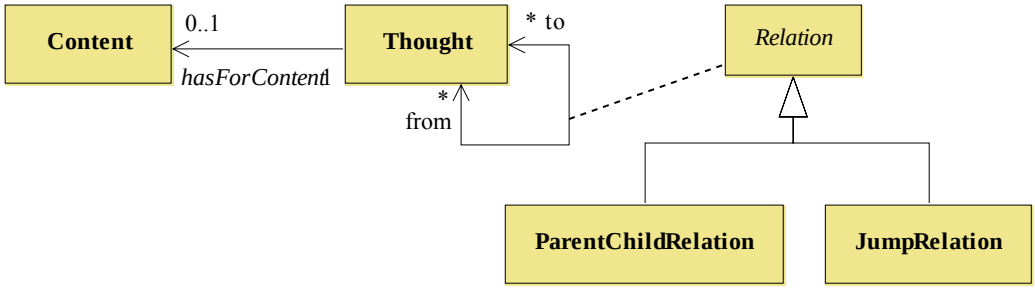


Figure 4.7 : Rétro-conception de *TheBrain*® (diagramme de classe UML)

La contre partie de la simplicité du modèle intertextuel réside dans un certain nombre de limites. La première concerne la gestion des contenus : ceux-ci sont gérés de manière monolithique, c'est-à-dire sans notion de fragment. La seconde concerne les traitements appliqués au réseau. Ceux-ci se bornent à l'affichage, pour un nœud sélectionné, de ses voisins (parents, enfants, frères, sauts). En effet, il serait difficile d'obtenir des affichages plus synthétiques puisque aucun contrôle de cohérence n'est appliqué au réseau. Il risquerait d'y avoir, par exemple, des paradoxes (existence d'un cycle) dans les relations parents/enfants.

TheBrain® (dans sa version « entreprise ») permet l'écriture collaborative d'un réseau. Cependant, l'espace partagé est ici un espace de consensus et non de débat. Nous sommes donc loin d'une instrumentation de l'intersubjectivité.

3. Bilan provisoire

Suite à l'étude théorique réalisée dans les chapitres précédents, nous avons vu dans celui-ci que les enquêtes effectuées dans les bibliothèques traditionnelles corroboraient la nécessité de l'intertextualité et de l'intersubjectivité. Nous nous sommes alors intéressé à l'instrumentation de ces deux notions dans sept systèmes hypermédias (issus du monde académique ou de celui de l'industrie) qui nous ont semblé représentatifs. Il en ressort que même si l'intertextualité est un fondement de l'hypermédia, les modèles proposés sont rarement utilisables pour le cas qui nous occupe. Plus encore, la question de l'intersubjectivité reste encore largement à explorer, ne serait-ce que dans sa forme la plus simple : l'expression en un même lieu de plusieurs points de vue.

Dans de telles circonstances, l'instrumentation que nous cherchons de l'intertextualité et de l'intersubjectivité passe par la définition d'un nouveau modèle hypermédia. C'est ce que nous tâcherons de faire dans notre deuxième partie.

2ème partie : Modèle proposé

Chapitre 5. Traces et publication*

The owner of the memex, let us say, is interested in the origin and properties of the bow and arrow. [..] First he runs through an encyclopedia, finds an interesting but sketchy article, leaves it projected. Next, in a history, he finds another pertinent item, and ties the two together. Thus he goes, building a trail of many items. Occasionally he inserts a comment of his own, either linking it into the main trail or joining it by a side trail to a particular item. When it becomes evident that the elastic properties of available materials had a great deal to do with the bow, he branches off on a side trail which takes him through textbooks on elasticity and tables of physical constants. He inserts a page of longhand analysis of his own. Thus he builds a trail of his interest through the maze of materials available to him.

And his trails do not fade. Several years later, his talk with a friend turns to the queer ways in which a people resist innovations, even of vital interest. He has an example, in the fact that the outraged Europeans still failed to adopt the Turkish bow. In fact he has a trail on it. [..] It is an interesting trail, pertinent to the discussion. So he sets a reproducer in action, photographs the whole trail out, and passes it to his friend for insertion in his own memex, there to be linked into the more general trail.

Vannevar Bush, As we may think, 7.

---

* Ce chapitre fait l'objet d'un article de recherche à paraître dans les Suppléments au Bulletin de Correspondance Hellénique [Benel03].

Les notions de "trace"et de "publication" ont été par ailleurs introduites lors d'une conférence à la "European Conference on Digital Libraries" [BenelEtAl01a] et dans un article de vulgarisation [BenelEtAl01b].

1. Notion de trace

a. Principe

Le sens d'un document n'est pas dans sa structure (quelle qu'elle soit) mais dans les conditions de son utilisation et donc dans les conditions de sa lecture. Autrement dit, il est impossible de trouver *le* sens d'un document mais plutôt *des* sens. Toute lecture est donc une réécriture. Un système adéquat devrait permettre de garder trace des parcours d'interprétation des lecteurs. Vannevar Bush, visionnaire de ce qui allait devenir l'hypertexte, décrivait d'ailleurs en 1945 un modèle basé sur des traces (ou pistes) de pensée [Bush45]. Cette métaphore de la « trace » est particulièrement riche et mérite que l'on s'y attarde un peu.

Tout d'abord, une trace, c'est la forme qui demeure quand le contenu a disparu. En effet, comme nous l'avons vu, l'ordinateur ne traite que de symboles dénués de signification. C'est à l'utilisateur qu'incombe la difficile tâche de donner un contenu[48] à ce qui n'est que formel. Ensuite, une trace est rarement seule, elle fait partie d'une piste. On retrouve ici la notion de contexte, c'est-à-dire d'autres traces reliées à elle. Enfin, on peut « suivre quelqu'un à la trace » mais aussi « suivre les traces de quelqu'un ». Dit autrement, il existe deux approches complémentaires : interpréter les traces d'un autre selon notre point de vue, ou tenter par la pensée de se mettre « à sa place ». On retrouve ici la notion de situation.

Si les symboles en contexte peuvent être « rentrés » dans une machine, il n'en est pas de même pour les situations. Il est impensable de faire rentrer dans un ordinateur des personnes, des lieux, des cultures, bref le monde. Aussi nous considérerons plutôt des « indices de situation », c'est-à-dire des informations sur les différentes étapes de la « vie » d'un symbole contextualisé (date et auteur de la création d'un contexte, etc.).

Ainsi, notre approche théorique nous permet de proposer le patron de la « Trace » (cf. Figure 5.1). Pour un symbole donné, moyennant une heuristique (ex : locale, globale...), il est possible d'obtenir un contexte pointant vers un certain nombre

[48] Au sens philosophique.

d'autres symboles. Ce contexte est associé à des indices de situation. Notons qu'une trace peut être le résultat de l'agrégation d'autres traces, l'agrégation étant une mise en contexte particulière.

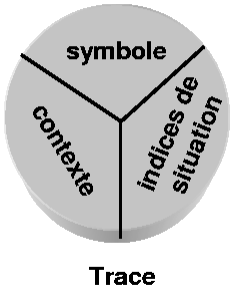


Figure 5.1 : Notion de trace

Dans notre prototype (appelé *Porphyre*) ces traces (cf. Figure 5.2) peuvent être des documents : fragments, sources ou notes. Elles peuvent également être des corpus de documents, des descriptions semi-formelles de ces corpus (à l'aide de graphes acycliques de descripteurs), ou encore des étapes de lecture. Les descriptions sont mises en contexte entre-elles suivant les relations de généralisation/spécialisation qui existent entre leurs graphes respectifs. Les étapes de lecture, pour leur part, sont mises en contexte par les relations de séquence indiquées dans des parcours de lecture.

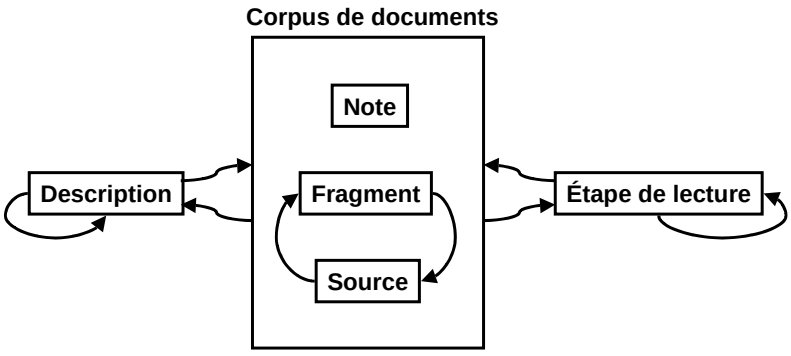


Figure 5.2 : Types de traces et liens contextuels

b. Architecture

Pour la gestion des documents et des corpus, nous utilisons un *serveur de contenu* (cf. Figure 5.3). Celui-ci stocke et diffuse les documents sources (images ou textes).

Il se distingue d'un serveur Web normal en permettant également d'extraire des fragments des documents sources, de « surligner » tous les fragments associés à un document source, ou encore, dans le cas où les documents sont des images, d'obtenir des zooms ou des vignettes. Les documents sources sont placés grâce à un serveur FTP (en mode passif pour passer à travers les pare-feux).

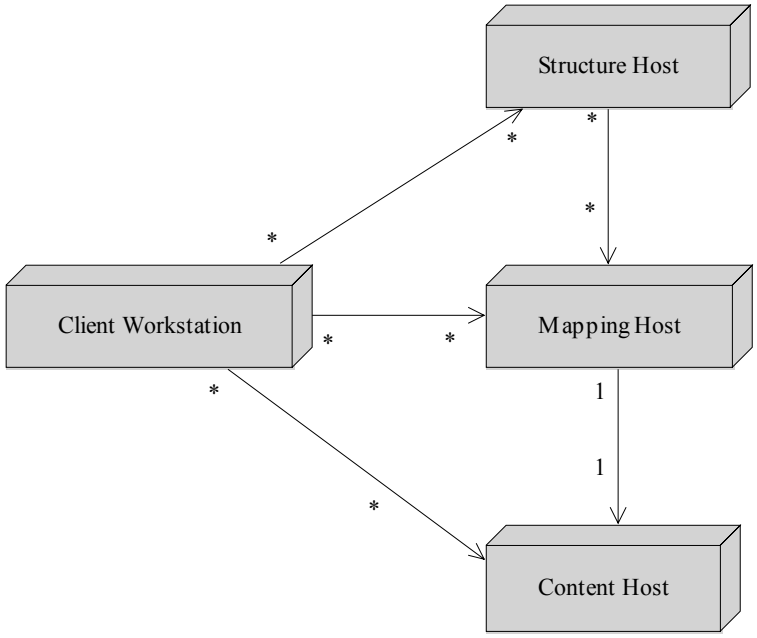


Figure 5.3 : Architecture multi-tiers du système *Porphyre* pour un accès natif (diagramme de déploiement UML)

Pour la gestion des traces semi-formelles, nous faisons appel à un *serveur de structure* (cf. Figure 5.3). Celui-ci stocke ces traces dans une base de données, et permet de les présenter dans un contexte susceptible d'intéresser le lecteur.

Les traces peuvent être consultées soit par un *client web* par le biais d'un *serveur passerelle* (cf. Figure 5.4) soit par le *client natif* (cf. Figure 5.3). Ce dernier destiné au lecteur expert lui permet d'effectuer une lecture-écriture des traces suivant la métaphore de l'annotation. Nous nous plaçons ainsi dans la prolifique descendance des travaux de Jacques Virbel et Bernard Stiegler autour des « Postes de Lecture Assistée par Ordinateur » [Stiegler00].

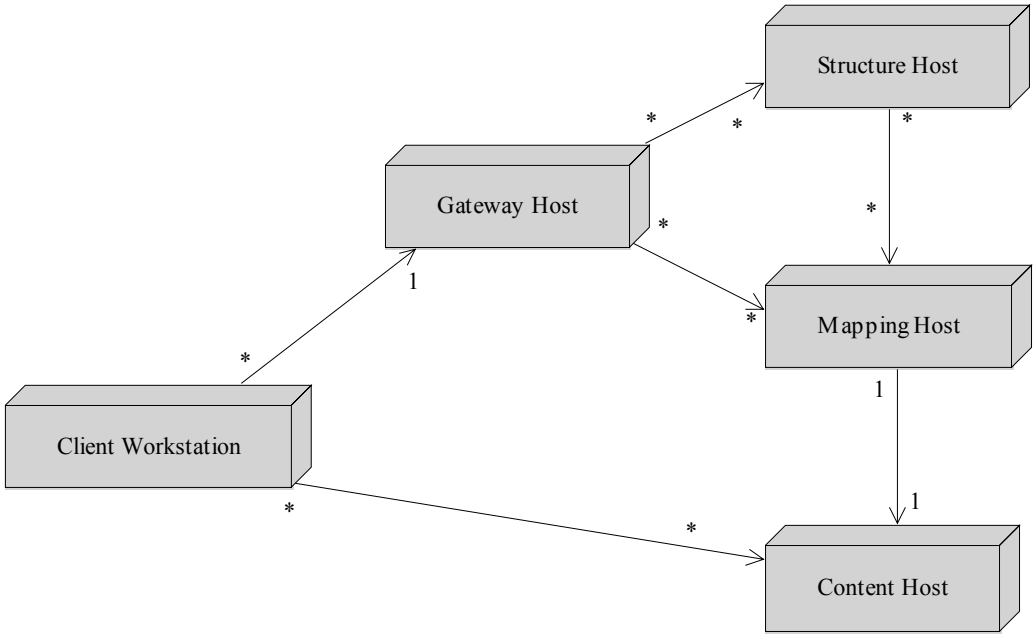


Figure 5.4 : Architecture multi-tiers du système *Porphyre* pour un accès par le Web (diagramme de déploiement UML).

Si nous n'avons pas parlé jusqu'à présent du *serveur de correspondance,* c'est que sa fonction est plus délicate à saisir : il construit l'URL[49] permettant d'obtenir un contenu. Son but est donc d'éviter une redondance des données (stocker à la fois les paramètres du contenu et l'URL qui en dépend).

Le fait que le serveur de correspondance soit distinct du serveur de contenu permet d'interfacer d'autres types de serveurs de contenu : versions antérieures de *Porphyre*, serveur *Transvision*®[50], Serveurs Web classique. Lorsqu'un de ces types de serveurs ne prend pas en charge une fonction donnée, la requête est alors simplifiée[51].

---

[49] *Uniform Resource Locator* [URL] : Adresse permettant de référencer toute ressource (statique ou dynamique) accessible sur Internet (entre autres sur le Web). Ex : http://porphyry.porphyry.org/image/getSource?id=BCH_120_1_492&maxWidth=800

[50] *Transvision*® : Système de gestion de bases d'images développé par l'équipe « images » de la Maison de l'Orient et de la Méditerranéen puis commercialisé pendant quelques années par la société « SGBI Entreprise SA» [TVML00]. Il s'agit du système utilisé pour la partie « contenu » du projet CEFAEL [Benel02b].

En étant distinct du serveur de contenu, le serveur de correspondance permet d'utiliser une adresse plus abstraite (comparable aux URN[52]) utilisable pour la comparaison de deux références provenant de deux serveurs de structure différents. Il devient alors possible d'associer à un serveur de contenu, plusieurs serveurs de structure, tout comme un serveur de structure peut l'être à plusieurs serveurs de contenu. Ainsi, plusieurs équipes de recherche peuvent travailler sur un même corpus, et réciproquement une équipe peut travailler sur plusieurs corpus à la fois.

2. Notions de publication, de communauté et d'édition

Il ne suffit pas de stocker des traces, encore faut-il les partager... La métaphore que nous choisirons est celle de la *publication* (cf. Figure 5.5). Cette « mise en public » est basée sur une relation de confiance concrétisée par une communauté. L'adhésion d'un expert à une communauté nécessite l'accord commun de l'expert et du représentant de la communauté. Membre de la communauté, l'expert peut lire les documents et les traces publiés par les autres membres. Dans une bibliothèque privée, il peut ajouter ses propres traces d'interprétation relatives à ses propres documents ou aux documents de la communauté.

---

[51] Par exemple, si l'on demande l'URL permettant d'extraire un fragment d'image avec redimensionnement sur un serveur Transvision®, l'URL construite tiendra compte de l'ordre de redimensionnement mais pas de celui d'extraction de fragment.

[52] *Uniform Resource Name* [URN] : Identifiant permanent d'une ressource sur Internet, indépendamment de sa localisation physique. Ex : « urn:hdl:cnri.dlib/august95 », « urn:doi:10.1000/1 », « urn:issn:1082-9873 »…

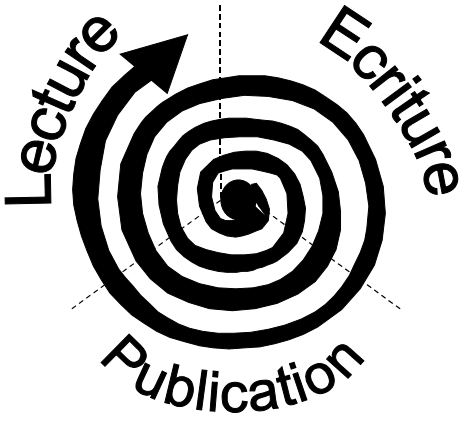


Figure 5.5 : La publication comme moteur de l'interprétation

Si l'expert le souhaite, il peut soumettre au représentant de la communauté une exportation totale ou partielle de ses traces et documents personnels pour les publier. Le responsable de la communauté peut alors choisir de les publier ou non, éventuellement en y apportant des modifications consenties par l'auteur.

La figure du « représentant de la communauté » permet de gagner en généralité. Il incombe ensuite à chaque communauté de décider quelle organisation elle mettra derrière : anarchie, consensus, majorité, comité de lecture, ou dictature !

Dans un souci d'ouverture du système, la soumission d'un ensemble de traces se fait en dehors du système (par l'intermédiaire d'un courriel par exemple). Les traces sont exportées par leur auteur dans un fichier XML, le fichier est transmis au représentant de la communauté qui les importe ensuite dans une zone brouillon où il peut les modifier. Le représentant peut enfin publier (au sens propre) les traces, c'est-à-dire les transférer dans la zone publique de la communauté.

3. Histoire d'une trace

Nous venons de voir de manière informelle les processus qui affectaient dans le temps un ensemble donné de trace. Notre but serait maintenant de capter l'historique de ces transformations.

Pour ce faire, nous devons tout d'abord conserver l'ensemble des symboles. Seules les relations contextuelles pourront être modifiées. Elles le seront par les actions suivantes : *création*, *suppression*, *archivage, purge*[53].

Ces actions feront passer les relations contextuelles par quatre états (cf. Figure 5.6) :

- *Brouillon* : présent dans l'espace du scripteur, mais pas encore dans celui du lecteur,
- *Publié* : présent dans les deux espaces,
- *Obsolète* : encore présent dans l'espace du lecteur, mais plus dans celui du scripteur,
- *Au pilon* : absent des deux espaces.

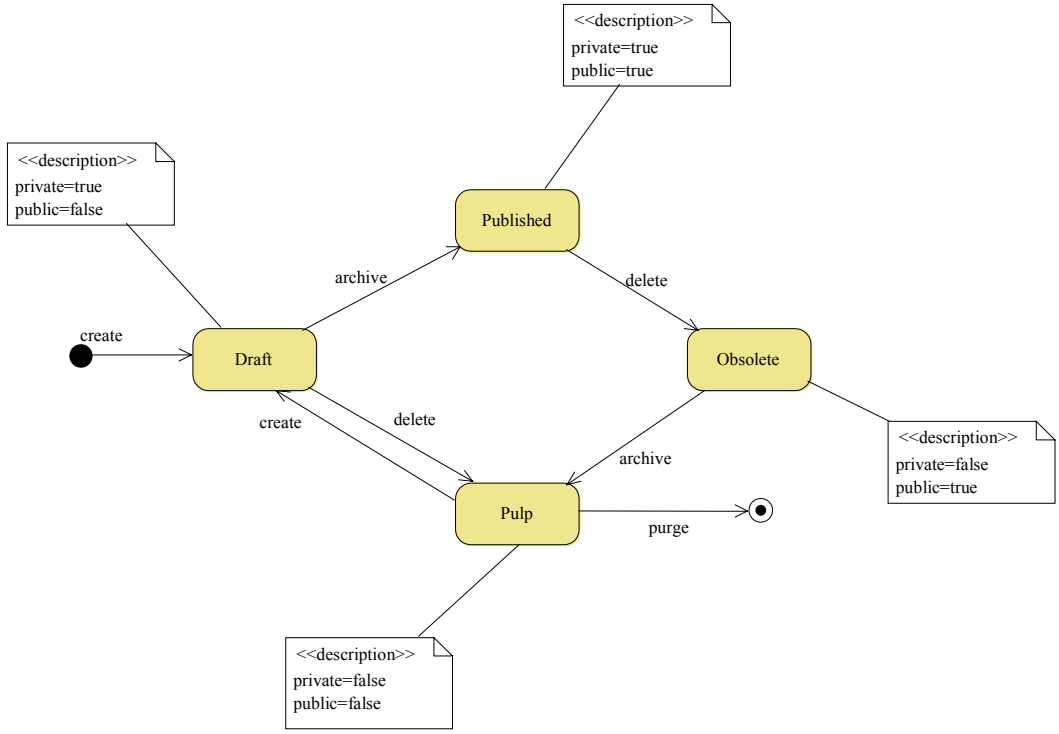


Figure 5.6 : « Cycle de vie » d'un contexte (diagramme d'état UML).

Notons que ce cycle de vie est basé sur la théorie de la connaissance exposée dans le chapitre 2. Ainsi, le parcours menant de l'état initial au *pilon* en passant par le

[53] Nous nous inspirons ici des serveurs de gestion de version (Ex : *Concurrent Version Server*, *Ms Visual Source Safe*, etc.).

brouillon, la *publication* et l'*obsolescence* reprend la théorie poppérienne de l'hypothèse : sa construction par un chercheur puis sa publication par la communauté, sa remise en question par un autre chercheur, et la publication de cette remise en question (réfutation). De même, la transition menant du *pilon* au *brouillon* reprend la théorie kühnienne selon laquelle les « acquis » scientifiques (donc les réfutations) peuvent être remis en question par une révolution des paradigmes. A ces transitions issues de la philosophie des sciences, nous en avons ajouté deux. Celle menant du *brouillon* au *pilon* rend compte de l'avortement de certaines hypothèses avant leur publication. Celle menant du *pilon* à l'état final (et donc à la destruction réelle) rend compte de la pratique des archivistes qui pour mieux conserver certaines choses doivent parfois en jeter d'autres.

Pour que l'historique soit complet, nous devons maintenant conserver l'ensemble des actions effectuées sur les relations contextuelles. Ainsi à chacune de ces relations sera associé un ensemble de situations[54] définies par trois questions :

- *Quoi* : nature de l'action (*création*, *suppression*, *archivage*),
- *Quand* : date et heure de l'action,
- *Qui* : nom et prénom du chercheur ayant effectué l'action.

Le modèle obtenu rend possible dans un premier temps certaines fonctionnalités simples mais intéressantes comme :

- la définition de deux espaces, un pour le scripteur et un pour le lecteur,
- l'obtention de l'historique d'une relation contextuelle,
- la possibilité d'annuler une suite erronée d'opérations (*undo/redo*).

Cependant, à terme, le but est de visualiser de manière synthétique l'activité dans le temps d'une communauté. Cet aspect sera développé dans le chapitre 12 comme perspective à nos travaux.

[54] Ou pour être plus précis d'*indices de situation*, puisque comme nous l'avons vu une situation ne peut être stockée dans un ordinateur.

Dans les chapitres suivants, nous détaillerons trois types de traces particuliers :

- les contenus documentaires,
- les réseaux de description,
- les parcours de lecture.

Chapitre 6. Contenus documentaires

Most of the memex contents are purchased on microfilm ready for insertion. Books of all sorts, pictures, current periodicals, newspapers, are thus obtained and dropped into place. Business correspondence takes the same path. And there is provision for direct entry. On the top of the memex is a transparent platen. On this are placed longhand notes, photographs, memoranda, all sorts of things. When one is in place, the depression of a lever causes it to be photographed onto the next blank space in a section of the memex film, dry photography being employed.

Vannevar Bush, As we may think, 6.

Dans le chapitre précédent, nous avons mis en place un modèle à base de traces. Dans ce chapitre-ci, nous définirons un premier type de trace que nous appellerons « *contenu documentaire* ».

Le lecteur pourrait s'étonner que l'on consacre un chapitre à un type de trace dont la gestion serait *a priori* plus du domaine de l'ingénierie que de celui de la recherche. Cependant, comme ces contenus documentaires serviront de support aux types de traces que nous verrons par la suite, et qu'aucun outil du commerce, à notre connaissance, n'assure l'intégralité des fonctions proposées[55], il ne nous semble pas superflu d'en faire une présentation détaillée.

[55] Le système Transvision®, déjà cité, bien que proche de ce que l'on souhaite, n'assure qu'une partie des fonctionnalités recherchées [TVML00].

1. Notions

a. Contenu documentaire

« Qu'est ce qu'un document ? » : la question est loin d'être naïve[56]. Prenons l'exemple d'une collection scientifique en ligne (par exemple une revue). Le document se situe-t-il au niveau du paragraphe et de l'illustration ? Du fac-similé de la page ? Du tome ? Du volume ? De la collection complète ? Nous nous abstiendrons ici de faire du document une définition même semi-formelle. Sera « document » ce qu'un individu considèrera comme « document ».

Par conséquent, comme primitive de notre système, nous ne prendrons pas le document, mais tout simplement la « granule » choisie pour le stockage et nous l'appellerons « un *contenu documentaire* ».

Pour être archivé, un contenu documentaire doit être aussi stable que possible. Par conséquent chaque nouvelle version d'un contenu documentaire fera l'objet d'un nouveau contenu documentaire. De même, la clef de référence d'un contenu documentaire ne pourra être modifiée.

b. Objet documentaire

A l'usage, il apparaît très vite que la seule notion de *contenu documentaire* n'est pas suffisante.

Par exemple, lors du projet de numérisation des collections de l'EFA, chaque tome, une fois massicoté, a été placé dans un chargeur pour être numérisé recto-verso. Ainsi, chaque fac-similé de page pouvait être référencé automatiquement par un couple d'entiers : le numéro d'ordre du « codex » numérisé et le rang de la page dans ce codex. Pour référencer et feuilleter convenablement les fac-similés (par la table des matières, des figures, etc.), notre équipe a dû proposer une nomenclature comprenant le nom

[56] Cette question fait d'ailleurs l'objet actuellement d'une rédaction collective au sein du Réseau Thématique Pluridisciplinaire « Documents et contenu : création, indexation, navigation » (CNRS) : <http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/04/13>.

abrégé de la collection, le numéro de volume, le numéro de tome, le type de page (préliminaire, foliotée, finale, planche) et le folio. Ainsi, comme le montre l'exemple de la figure 6.1, le 4ème fac-similé du 231ème codex numérisé correspondait en fait au 1001ème folio du 3ème tome du 121ème volume de la revue BCH.

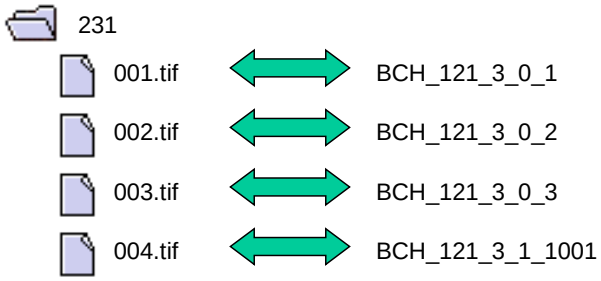


Figure 6.1 : Exemple de correspondance pour un contenu documentaire entre une référence automatique et un nom significatif.

Notons que l'obtention de cette nomenclature finale n'a été possible qu'au prix de l'abandon d'autres nomenclatures et donc au prix du changement (automatique mais long) du nom de tous les fac-similés numérisés. Si l'on refaisait aujourd'hui la numérisation, il serait sans doute préférable de distinguer pour un fac-similé sa référence automatique de son nom.

De manière plus générale, pour manipuler un contenu, il faudra lui donner un nom. Contrairement à la référence automatique, ce nom, résultat d'une interprétation, peut éventuellement être modifié. Nous nous trouvons donc en présence d'un autre niveau que nous appellerons « une *source* ».

Nous avons considéré jusqu'à maintenant les contenus documentaires comme des atomes[57], des éléments amorphes, sans structure. Or, par le seul fait de son inscription sur un support, l'élément documentaire *est* structuré. Dit autrement, l'élément documentaire, par sa structure interne, définit un ensemble de parties virtuellement adressables[58]. Par exemple, une image dans sa représentation matricielle définit virtuellement

[57] Au sens étymologique (indivisible).

[58] Nous reprenons ici la terminologie que nous avons définie (en nous inspirant entre autres de *Xanadu*) au sein du groupe de réflexion de l'ISDN sur les « documents multi-structurés ».

un très grand nombre de zones rectangulaires. La même image, dans une représentation fréquentielle, définit virtuellement un ensemble de version de l'image avec plus ou moins de détails. Lorsqu'une partie virtuellement adressable sera utilisée par un être humain, nous en garderons trace et l'appellerons « un *fragment* ».

Enfin, à la demande des utilisateurs, nous avons été amenés à définir les *notes de lecture* comme des éléments dynamiques qui à la différence des sources peuvent être modifiées sans créer de nouvelles versions. Notons que le caractère dynamique de leur contenu nous empêche de définir dessus des fragments[59].

Nous définissons la notion d'*objet documentaire* comme la généralisation des notions de « source », de « fragment » et de « note » (cf. Figure 6.2). Cet objet documentaire est caractérisé par le couple formé :

- d'un espace de nom (le numéro IP de son serveur de correspondance),
- et d'un nom, aussi significatif que possible (dans l'exemple de la Figure 6.1: « BCH_121_3_1_1001 »).

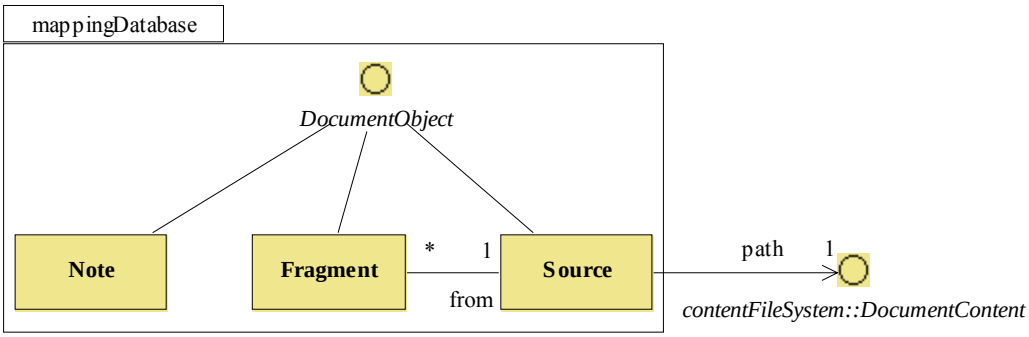


Figure 6.2 : Contenus documentaires (diagramme de classe UML)

Ce couple de valeur appelé « *localisation d'objet documentaire* » sera utilisé comme référence par les autres types de trace. On considèrera que deux traces font référence au même objet documentaire si et seulement si leur localisation d'objet documentaire est identique.

[59] En effet une étude portant sur le « balisage » de textes évolutifs mériterait sans doute une thèse à elle toute seule.

2. Traitements

a. Sur un objet documentaire isolé

Pour obtenir une objet documentaire dont on connaît la localisation (par exemple : « 134.214.105.147/BCH_121_3_1_1001 »), il faut s'adresser au serveur de correspondance de la source (« 134.214.105.147 ») en lui fournissant le nom de la source (« BCH_121_3_1_1001 »). Celui-ci nous renvoie un objet de la classe abstraite *DocumentObject*, instancié en fonction du type de serveur de contenu (ici, la version 2003 du serveur de contenu de *Porphyre*). Cet objet comporte un certain nombre de méthodes permettant entre autres d'obtenir l'URL de visualisation (en fonction d'une largeur maximale donnée) et celle de sa vignette.

La mention d'une largeur maximale permet pour des contenus documentaires de type image, archivés à très haute définition[60], d'obtenir des vues redimensionnées à la baisse en fonction des besoins[61].

Si notre objet documentaire est une image source, nous obtiendrons des URL du type :

- « http://contentserver.porphyry.org/Image/getThumbnail?file=231/4 » pour sa vignette (cf. Figure 6.3a),
- « http://contentserver.porphyry.org/Image/getSource?file=231/4&max=640 » pour la vue réduite à *640* pixels de largeur maximum (cf. Figure 6.3b).

S'il s'agit d'un fragment d'image, nous obtiendrons une URL du type :

- « http://contentserver.porphyry.org/Image/getFragment?file=231/4&coord=1000+1100+700+400&max=600 » pour la vue obtenue par extraction de la zone ayant pour coin supérieur gauche, le point de coordonnée cartésienne (*1000,1100*), pour largeur *700* et pour hauteur *400*. La vue après extraction est réduite à *640* pixels de largeur maximum (cf. Figure 6.3c).

[60] Et souvent compressés sans pertes (par exemple en TIFF).

[61] Et compressées avec pertes – par exemple en JPEG – pour plus de fluidité sur le réseau.

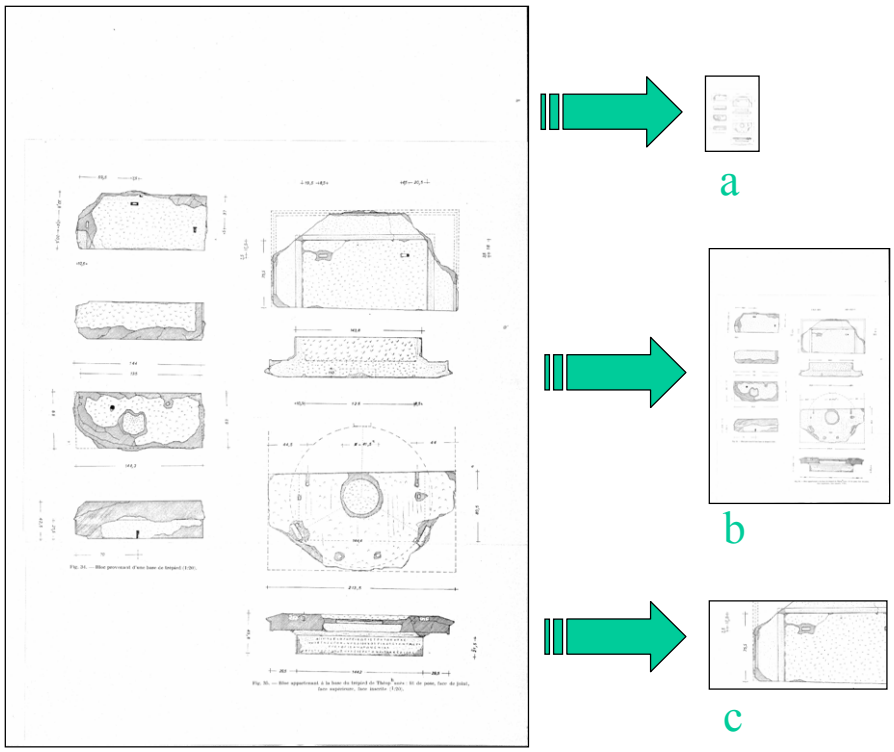


Figure 6.3 : A partir d'un même contenu documentaire : a. une vignette, b. une vue de la source, c. une vue d'un fragment.

Ces URL, et d'autres permettant de gérer le texte intégral, font appel à des scripts PHP du serveur de contenu de *Porphyre*. Nous invitons le lecteur intéressé par l'implémentation optimisée qui en a été faite de se reporter au mémoire CNAM en cours de rédaction de Régine Tribollet [Tribollet03].

b. Sur un contexte de lecture

Dans l'approche intertextuelle qui est la nôtre, l'objet documentaire ne peut se comprendre que dans un ensemble. Nous appellerons cet ensemble « un *contexte de lecture* ». Or, il serait illusoire de penser que gérer un tel contexte se ramène à mettre bout à bout plusieurs objets documentaires. Il s'agit au contraire de trouver des métaphores formelles et visuelles à la « sélection » de sens qui s'effectue selon François Rastier entre deux textes[62] lus en vis-à-vis.

[62] Au sens large (cf. chapitre 3) : texte intégral, photographie, diagramme…

La requête au serveur de correspondance ne porte donc plus sur un objet documentaire isolé mais sur un contexte de lecture. Les URL construites pour chaque objet peuvent alors tenir compte de ce contexte de lecture.

Dans la version 2003 du système *Porphyre*, un premier traitement du contexte de lecture a été mis en place. Il vise à matérialiser dans une source la relation qu'elle entretient avec ses fragments quand ils sont lus en contexte[63]. Ceci est valable aussi bien pour des contenus textuels que graphiques (cf. Figure 6.4).

L'URL du document source, avec encadrement des zones appartenant à ses fragments est alors de la forme (cf. [Tribollet03]) :

- « http://contentserver.porphyry.org/Image/getSource?file=231/10&coord=600+450+150+100;760+400+200+100&max=640 ».

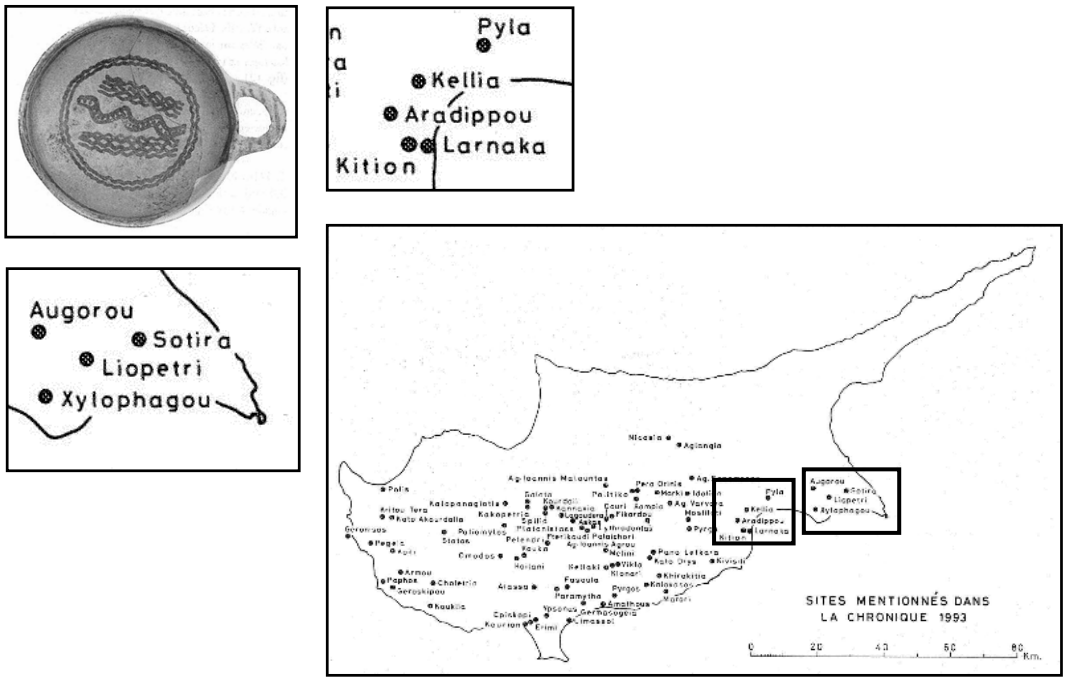


Figure 6.4 : Fonction d'encadrement automatique dans un contexte de lecture comprenant des fragments et leur source.

Nous envisageons de mettre en place par la suite d'autres traitements portant sur les contextes de lecture. Quand un de ces derniers comporte des objets textuels, il pour-

[63] Nous nous inspirons ici de la visualisation dans *Xanadu*® des liens de citation [Nelson99].

rait être intéressant, par exemple, de distinguer graphiquement les termes propres à un objet de ceux que l'on retrouve dans plusieurs. Dans le même ordre d'idée mais avec un aspect plus statistique, l'utilisation de l'incontournable « *tf.idf* »[64] permettrait de faire ressortir les termes à la fois fréquents dans un objet documentaire et rares dans le contexte de lecture.

[64] *tf.idf* (de l'anglais : « term frequency, inverse document frequency ») : Variable statistique couramment utilisée en recherche d'information pour extraire des termes présents dans le texte intégral d'un document de telle sorte qu'ils soient les plus discriminants possible par rapport au corpus.

Chapitre 7. Réseaux de description*

> *The real heart of the matter of selection, however, goes deeper than a lag in the adoption of mechanisms by libraries or a lack of development of devices for their use. Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome.*
> Vannevar Bush, As we may think, 6.

Dans le chapitre précédent, nous avons défini la notion d'objet documentaire. Dans ce chapitre-ci, nous verrons comment en décrivant de manière structurée ces objets, nous les mettrons en contexte les uns avec les autres. Suivant la nature des objets documentaires, le résultat de cette mise en contexte pourra être interprété comme un document (au sens traditionnel), un corpus de documents, ou encore comme un document « sur mesure ». Autrement dit, nous ne ferons pas de différence formelle entre :

- la description d'une section parmi un document (rédaction),
- celle d'un document parmi une collection (indexation),
- ou encore celle d'un fragment parmi un dossier bibliographique (annotation).

Dans un premier temps, nous étudierons comment la question de la description de documents (de manière structurée) est ordinairement traitée. Dans un deuxième temps, nous proposerons un modèle basé sur la notion de « point de vue ». Enfin, dans un dernier temps, nous verrons comment il est possible d'offrir des outils basés sur notre modèle permettant « d'arpenter » l'espace documentaire.

---

* Des parties de ce chapitre ont fait l'objet des conférences suivantes :

- l'International Symposium on Methodologies for Intelligent Systems [BenelEtAl02],
- le Colloque International Francophone sur l'Ecrit et le Document [BenelEtAl00b],
- le Congrès INFORSID [BenelEtAl00a],

1. Décrire de manière structurée des objets documentaires

a. Des arbres qui cachent… la bibliothèque

L'exergue de ce chapitre rappelle que l'organisation traditionnelle des bibliothèques est basée sur l'idée que chaque livre traiterait d'un sujet unique, sujet qui lui-même serait situé sans ambiguïté possible dans une hiérarchie universelle. L'auteur (Vannevar Bush) dénonce ici le caractère artificiel et inutilisable de ce modèle arborescent instauré dans les bibliothèques par Dewey (1859-1952).

Les racines du problème remontent sans doute à Aristote (IV° s. av. J.-C.). Celui-ci nous déclare en effet dans les *Catégories* :

> Quand une chose est attribuée à une autre comme à son sujet, tout ce qui est affirmé du prédicat devra être aussi affirmé du sujet […].
>
> Si les genres sont différents et non subordonnés les uns aux autres, leurs différences seront elles-mêmes autres spécifiquement […].
>
> Par contre, dans les genres subordonnés les uns aux autres, rien n'empêche que leurs différences soient les mêmes, car les genres plus élevés sont prédicats des genres moins élevés, de sorte que toutes les différences du prédicat seront aussi des différences du sujet. [Aristote-300, Catégories 3]

Ces trois principes sont interprétés par Porphyre[65] (III-IV° s. ap. J.-C.) comme définissant un arbre dans lequel à chaque nœud un genre (*genus*) se divise en sous-genres en fonction d'une différence (*differentia*). Chacun de ces trois principes peut donc se lire de manière appliquée dans « l'arbre de Porphyre » (cf. Figure 7.1) :

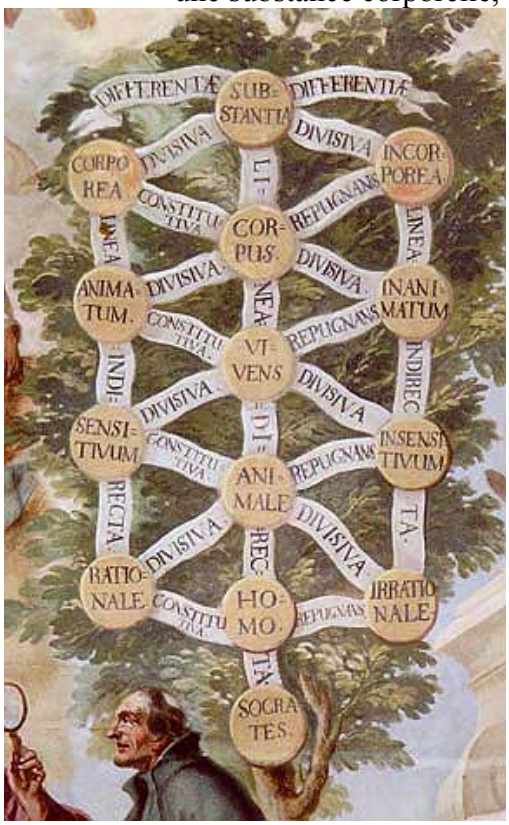
1. Les hommes sont des animaux. Les animaux sont sensibles. Donc les hommes sont sensibles.
2. « Rationnel » est une propriété « d'Homme » qui ne subordonne pas « Végétal ». Donc, « Rationnel » n'est pas une propriété de « Végétal ».
3. L'Homme est :
   un animal rationnel,

- l'Escuela interlatina de altos estudios en linguistica aplicada [BenelEtCalabretto00],

- le Chapitre français de l'ISKO [BenelEtAl99].

[65] dans son introduction aux *Catégories.*

un être vivant sensible et rationnel,
une matière animée, sensible et rationnelle,
une substance corporelle, animée, sensible et rationnelle.



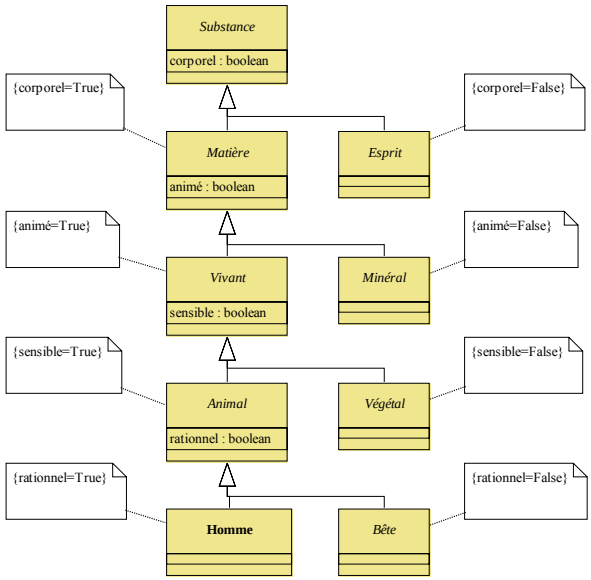


Figure 7.1 : « L'arbre de Porphyre » : Fresque du XVIII[e] s.[66] et diagramme de classe UML.

On comprend sans peine qu'une organisation des connaissances d'une telle esthétique et d'une telle efficacité soit devenue le parangon de la pensée occidentale. Cependant, si cette méthode est tout à fait valide pour parler de classes, elle ne devrait en aucun cas être utilisée pour des instances, celles-ci pouvant souvent être placées dans plusieurs classes contradictoires. En effet, on pense tout de suite au célèbre exemple de Nixon potentiellement pacifiste en tant que quaker et belliciste en tant que républicain.

[66] Détail d'une fresque du monastère de Schussenried (Allemagne) ; Artiste : Hermann, 1757 ; Photographe : J. Garrett, 2000 ; Disponible sur Internet : <http://www.library.northwestern.edu/collections/garrett>

b. Alternatives

La méthode arborescente étant inutilisable pour classer des instances, *a fortiori* elle l'est également pour des livres dont la description pourra contenir des classes[67], des instances et des liens entre instances. Aussi, depuis longtemps, des alternatives au modèle d'indexation hiérarchique ont été proposées[68]. La plus connue en sciences de l'information est celle de Ranganathan (1872-1972) appelée aussi « indexation par facettes », mais les plus ambitieuses sont sans doute celles basés sur les graphes conceptuels de John F. Sowa [MechkourEtAl95, Martin96, Genest00]. Dans ces dernières, chaque document est décrit par un graphe (distinct) comprenant des objets (éventuellement génériques) et des liens entre ces objets (cf. Figure 7.2). Ces objets et ces liens, sont des instances d'un modèle du domaine[69] (cf. Figure 7.3).

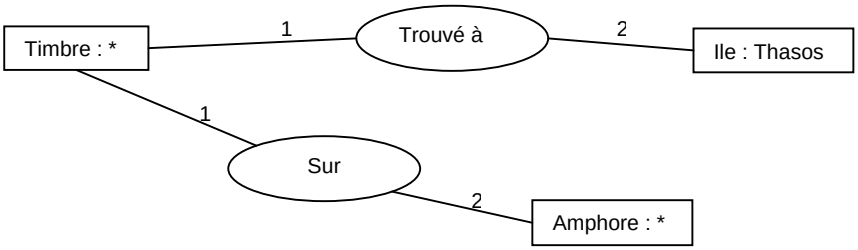


Figure 7.2 : Indexation à l'aide des graphes conceptuels d'une monographie traitant des timbres amphoriques thasiens.

[67] Ou tout au moins des objets génériques.

[68] Notons que la même critique a été faite pour l'organisation des objets documentaires dans un document, en réaction à la suprématie du format XML dans ce domaine. Voir par exemple [RenearEtAl96] et [Nelson97] ou encore le modèle proposé par notre groupe de réflexion au sein de l'ISDN pour les documents à structure multiple.

[69] Notons que cette séparation entre le modèle du domaine et les graphes d'indexation reprend celle existant dans les bibliothèques entre les « autorités » et les fiches d'indexation.

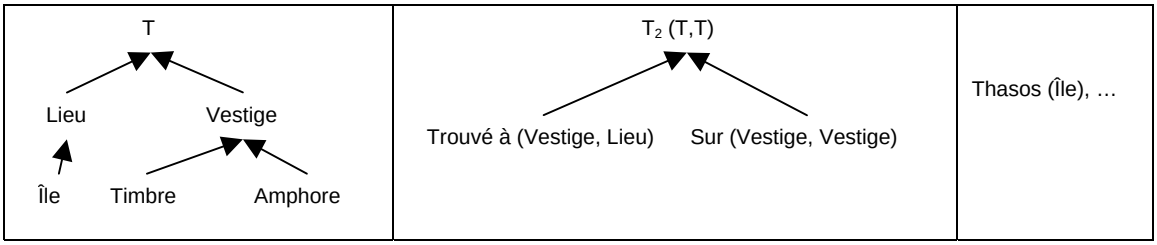


Figure 7.3 : Modèle du domaine nécessaire à l'indexation de la figure précédente : concepts, relations et instances.

Malheureusement, malgré leur puissance d'expression, ces alternatives sont, comme nous allons le voir, assez peu adaptées au cas qui nous occupe : celui d'une modélisation *dynamique* effectuée *par les usagers* de la bibliothèque.

c. Réfutation

Prenons un exemple en archéologie. Philippe Bruneau [Bruneau76], en réaction aux premières « banques de données archéologiques », faisait remarquer l'impossibilité de décrire « objectivement » une photographie du type de la Figure 7.4. Etait-on en présence de la représentation d'une mosaïque noire sur fond blanc ou blanche sur fond noir ? Plus grave encore, l'auteur nous faisait même douter du bien fondé d'une telle typologie.

Figure 7.4 : Mosaïque noire sur fond blanc ou blanche sur fond noir ? [Bruneau76]

Dans un tel cas, nous devrons disposer d'un modèle permettant d'exprimer qu'un premier point de vue affirme qu'il s'agit d'une mosaïque noire sur blanc, qu'un second affirme l'inverse, et qu'un troisième propose une typologie toute autre.

Les deux premiers points de vue étant contradictoires, notre « modèle de connaissance » devra être beaucoup plus permissif que la normale :

- Les structures seront non hiérarchiques (graphes orientés acycliques),
- Il n'y aura pas de négation (donc pas de principe de tiers-exclu, ni de principe de non-contradiction),
- Les points de vue ne seront pas dépendants les uns des autres, si ce n'est par l'intermédiaire des corpus décrits.

Du fait que le troisième des points de vue remette en cause la typologie utilisée dans les deux premiers, nous ne pourrons plus considérer qu'il existe *un* modèle *fixe* du domaine, mais plutôt *des* modèles *hypothétiques* et *transitoires*, évoluant de pair avec leurs instances. La séparation des classes et des instances en deux espaces apparaît par conséquent inutile. De manière plus générale, l'aspect dynamique de la modélisation empêchera un typage trop fort des primitives[70].

Les descriptions n'étant plus normées, il sera impossible de connaître *a priori* leur forme. Les interactions homme-machine ne devront donc pas suivre le modèle question-réponse mais plutôt celui de la navigation. La recherche de documents se fera donc de manière itérative et ira de pair avec une découverte de la structure du corpus. Dans une telle approche, la description du document sera un sous-graphe de la description du corpus[71]. En ce sens, nous nous rapprocherons un peu des techniques qui visent à agréger des graphes disjoints afin de donner une vision d'ensemble [Chalendar97, PredigerEtWille99, BurrowEtEklund94, EklundEtCole02].

Enfin, le fait que les experts ne soient pas des professionnels de la modélisation, nous encourage à proposer un modèle dont l'utilisation pour des descriptions simples sera assez intuitive, et dans lequel, il sera possible, moyennant une formation, d'établir

[70] Que l'informaticien qui n'a jamais changé, au cours d'un processus de modélisation, une instance en classe, ou une classe concrète en classe abstraite, nous jette la première pierre…

[71] Contrairement aux travaux cités précédemment où chaque document était décrit par un graphe distinct des autres.

des descriptions plus précises. Nous donnerons plus loin une esquisse du contenu d'une telle formation (cf. Chapitre 11).

2. Un modèle à base de points de vue

Notre modèle appelé *réseau de description* se présente sous la forme d'un graphe orienté acyclique (cf. Figure 7.5). Les nœuds sont appelés des « *descripteurs* » et les arcs des « *spécialisations* ». Un arc orienté entre les descripteurs *A* et *B* se lit : « tout objet documentaire décrit par *B* l'est aussi par *A* ».

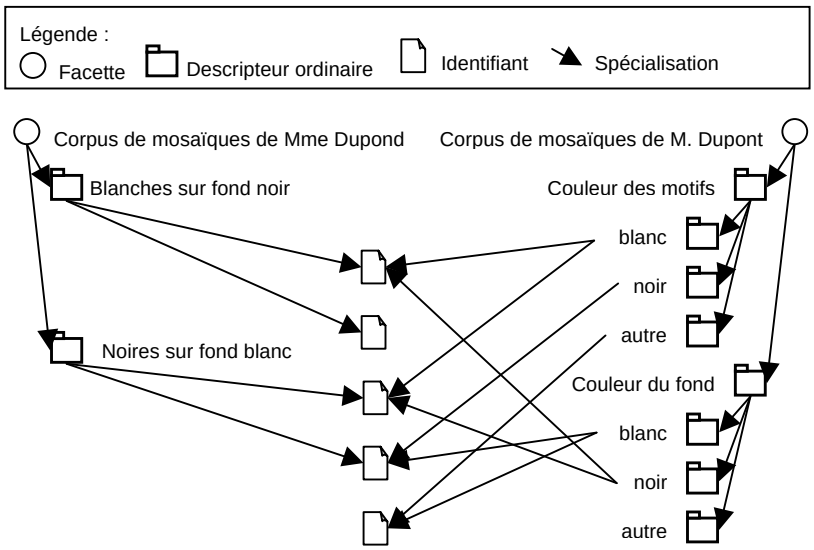


Figure 7.5 : Exemple de réseau de description

Il est important de mentionner que seul les nœuds et les arcs sont utiles pour le système. Mais de sorte que les usagers puissent interpréter le réseau, nous associons à chaque nœud une étiquette et à chaque arc son historique (cf. Chapitre 5). Du moment que la définition formelle des arcs est respectée, l'usager est libre d'utiliser ces arcs pour modéliser (cf. Chapitre 11) des taxinomies, des méréonymies, des instanciations…

Parmi les descripteurs (cf. Figure 7.6), certains ne sont pas généralisables : on les appelle des « facettes ». Chacun correspond à un point de vue indépendant. D'autres ne sont pas spécialisables, on les appelle des « *identifiants* ». Chacun fait référence à un objet documentaire unique.

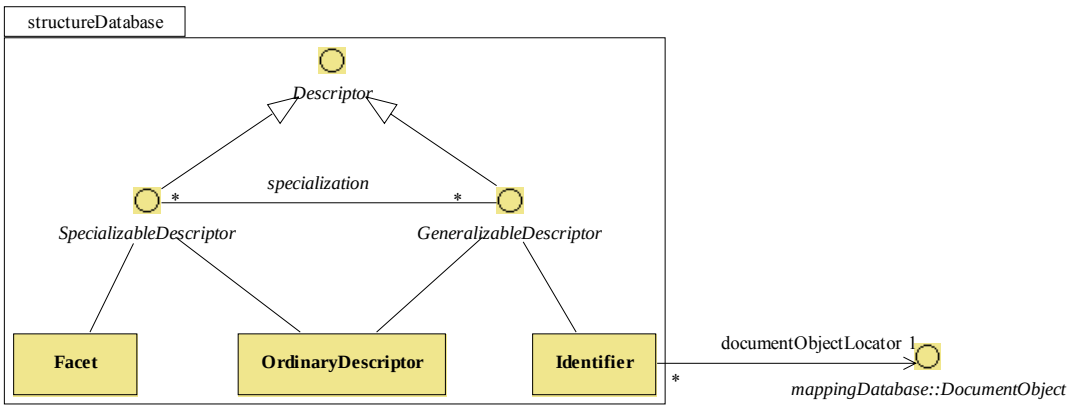


Figure 7.6 : Réseau de description (diagramme de classe UML)

Plus formellement, nous avons affaire aux ensembles suivants : *Descriptor, SpecializableDescriptor, GeneralizableDescriptor, Identifier, Facet, DocumentObject*.

Ces ensembles satisfont axiomatiquement les relations suivantes :

$Descriptor = SpecializableDescriptor \cup GeneralizableDescriptor$

$Identifier \subset GeneralizableDescriptor$

$Facet \subset SpecializableDescriptor$

Nous allons maintenant définir par des spécifications algébriques[72] :

- le schéma des données à stocker (primitives),
- les contraintes supplémentaires que ces données doivent respecter (contraintes),
- les requêtes complexes qui seront effectués sur ces données (définitions).

Primitive : $specialization(\_,\_) : SpecializableDescriptor \times GeneralizableDescriptor \rightarrow Boolean$

Primitive : $\_.getDOI : Identifier \rightarrow DocumentObject$

Fonction permettant d'obtenir le document correspondant à un identifiant donné.

[72] Voir en annexe l'explication de la syntaxe utilisée.

Note: le fait que ce soit une fonction et non une relation traduit une contrainte d'intégrité fonctionnelle sur les données.

Définition : *describes*(_,_) : *Descriptor* × *GeneralizableDescriptor* → *Boolean*
Relation construite de sorte qu'elle soit réflexive et qu'elle constitue la fermeture transitive de la relation « *spécialization* ».

$describes(x,x)$

$describes(x,y) \leftarrow specialization(z,y) \wedge describes(x,z)$

Note : cet algorithme est ascendant, ce qui constitue une optimisation pour un usage « normal » du réseau de description (un descripteur ayant souvent des centaines d'enfants et juste quelques parents).

Contrainte : « Acyclicité »
Aucun cycle ne doit exister dans le réseau de description.

$\perp \leftarrow specialization(x,y) \wedge describes(y,x)$

Contrainte : « Enracinement »
Un descripteur ne doit pas appartenir à plusieurs facettes.

$\perp \leftarrow f_1 \in Facet \wedge f_2 \in Facet \wedge f_1 \neq f_2 \wedge describes(f_1,x) \wedge describes(f_2,x)$

3. Arpenter l'espace documentaire

Chaque structure créée par un expert est perçue comme la trace d'un *point de vue*, d'une théorie portant sur une partie de la discipline. En effet, structurer son corpus documentaire, c'est structurer son objet d'étude, sa discipline.

Dès lors que des points de vue sont publiés (au sens de « rendus publics ») par des experts, tous ceux qui le souhaitent peuvent les voir superposés au corpus. Toutefois, si

nous avons choisi de laisser chaque expert exprimer son point de vue, ce n'est pas pour glorifier une subjectivité débridée, mais pour autoriser le débat, pour viser l'intersubjectivité. Il s'agit donc de pouvoir comparer entre eux ces points de vue.

Le mécanisme que nous avons offert aux utilisateurs est un filtre de graphes. Il permet par induction totalisante de trouver des rapports entre descripteurs, non-dits au niveau des modèles, mais apparaissant dans leurs usages. Pour reprendre notre exemple de typologies de mosaïques, le système nous indiquerait que lorsque tel auteur décrit les mosaïques comme blanche sur fond noir, un autre les décrit « toujours » (ou « parfois », ou « jamais ») comme des mosaïques noires sur fond blanc.

Notons qu'en filtrant les informations, nous en réduisons la charge cognitive. Nous tenons donc compte de l'une des préoccupations centrale des systèmes interactifs de recherche d'information [Hearst99]. En effet, vu la taille importante que prendra le réseau de description au fur et à mesure de son utilisation, il est crucial de n'en présenter à un moment donné qu'une partie. En fournissant ainsi un moyen à l'usager de s'orienter efficacement, nous espérons l'aider à arpenter l'espace documentaire.

a. Filtrer des descripteurs « à plat »

Gerard Salton, à la fin des années soixante [Salton68], a défini un modèle pour la recherche d'information basé sur la théorie des ensembles. Ce modèle considère un ensemble des documents et un ensemble des « descripteurs ». Ainsi on peut tracer les graphes d'inclusion de corpus (cf. Figure 7.7) et de conjonction de requêtes (cf. Figure 7.8).

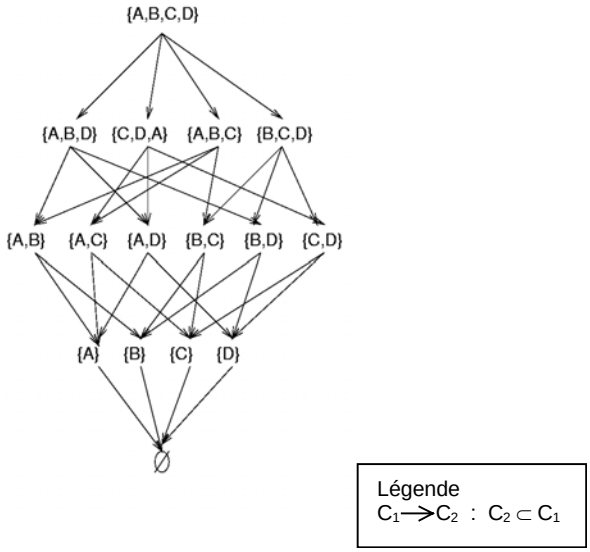


Figure 7.7 : Structure en treillis de l'espace des documents

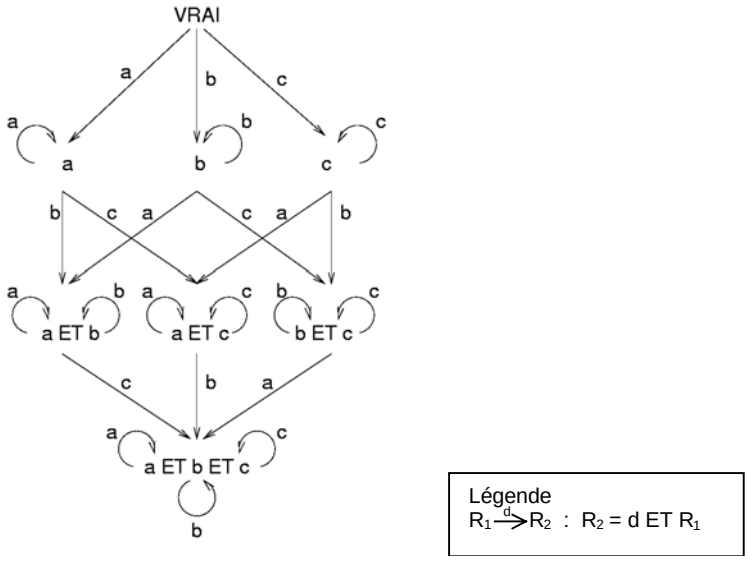


Figure 7.8 : Structure en treillis de l'espace des descripteurs

Il devient ensuite possible de déduire de la correspondance entre documents et descripteurs (cf. Tableau 7.1), la correspondance entre requêtes et corpus (cf. Tableau 7.2). De là, on remarque que certains corpus ne peuvent être obtenus par aucune requête (ex : {*B*,*C*}) et que le même corpus peut être obtenu par différentes requêtes (ex : la requête *a ET b* et la requête *b*). Ces résultats tout aussi connus qu'ils soient, ont été fort peu utilisés comme support des interactions homme-machines.

		Descripteurs		
		a	b	c
Documents	A	X		X
	B	X	X	
	C			X
	D	X		

Tableau 7.1 : Exemple de correspondance entre des documents et des descripteurs

VRAI	{*A*,*B*,*C*,*D*}
a	{*A*,*B*,*D*}
b	{*B*}
c	{*A*,*C*}
a ET b	{*B*}
b ET c	∅
c ET a	{*A*}
a ET b ET c	∅

Tableau 7.2 : Correspondance entre les requêtes et les corpus de documents (calculée à partir de l'exemple du tableau 7.1)

Claudio Carpineto et Giovanni Romano [CarpinetoEtRomano94] les ont utilisés en enlevant du graphe d'inclusion des corpus les corpus inaccessibles, obtenant ainsi un diagramme statique de généralisation/spécialisation des classes de document (cf. Figure 7.9).

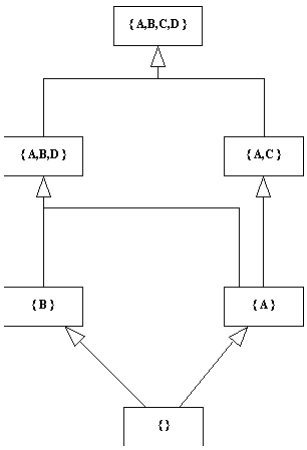


Figure 7.9 : Diagramme de classe (notation UML) dérivé de l'espace des documents

Dans notre approche, nous fusionnons dans le graphe des requêtes, celles qui décrivent le même corpus. Nous obtenons ainsi un diagramme d'état (cf. Figure 7.10) dans lequel les états correspondent à des corpus et les transitions à des requêtes élémentaires. Ces requêtes à un seul descripteur sur des corpus intermédiaires peuvent être vus comme l'ajout d'un descripteur à la requête globale : il s'agit d'une manière « d'affiner la requête » (en anglais : « query refining »).

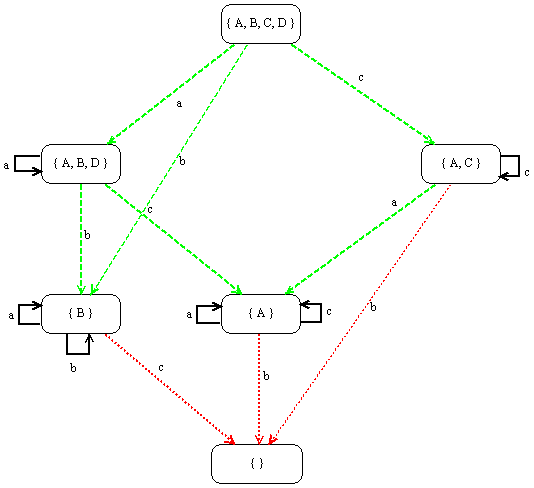


Figure 7.10 : Diagramme d'état (notation UML) dérivé de l'espace des descripteurs

Si nous reprenons le diagramme d'état précédent, dans un corpus donné chaque descripteur peut être dit :

- *impossible* : s'il mène du corpus actuel au corpus vide (ex : le descripteur *c* dans l'état {*B*}),
- *connu* : s'il « boucle » sur le corpus actuel (ex : le descripteur *a* dans l'état {*B*}, ainsi que le descripteur *b* dans l'état {*B*}),
- *possible* : dans les autres cas.

b. Filtrer des réseaux de descripteurs

Des trois états du descripteur définis précédemment, reste à voir comment définir un filtre pour le réseau. De manière informelle notre idée est la suivante : parcourir le réseau en descendant récursivement à chaque descripteur *connu* et s'arrêter aux descripteurs *possibles* ou *impossibles.* Soyons maintenant un peu plus précis[73]…

[73] Voir en annexe pour l'explication de la syntaxe utilisée dans les spécifications algébriques.

Définition : $\_.getCorpus : Descriptor \rightarrow DocumentObject^n$

Fonction permettant par récursivité d'obtenir l'ensemble des objets documentaires décrits directement ou indirectement par un descripteur donné.

Note : cet algorithme est déscendant car nous connaissons le descripteur et non l'objet documentaire.

$$x.getCorpus = \{y\} \leftarrow x \in Identifier \wedge x.getDOI = y$$

$$x.\ getCorpus = \{z \mid specialization(x,y) \wedge z \in y.\ getCorpus\} \leftarrow$$
$$x \in SpecializableDescriptor$$

Définition : $\_.getCorpus : Descriptor^n \rightarrow DocumentObject^n$

Généralisation de la fonction homonyme à une sélection de plusieurs descripteurs. L'intersection des corpus signifie qu'être décrit par une sélection de descripteurs revient à être décrit à la fois par chacun d'eux.

$$\{d_0,..d_n\}.getCorpus = d_0.getCorpus \cap ..d_n.getCorpus$$

Axiome

$$State = \{KNOWN, POSSIBLE, IMPOSSIBLE\}$$

Définition : $\_.getState(\_) : Descriptor \times DocumentObject^n \rightarrow State$

Fonction attribuant un état (*connu, possible, impossible*) à un descripteur pour un corpus *C* donné. Plus précisément, on dira que le descripteur est *connu* si le corpus du descripteur contient *C* et que *C* est non vide. Il sera *impossible* si le corpus du descripteur a une intersection vide avec *C.* Il sera *possible* dans les autres cas.

$$x.getState(C) = KNOWN \leftarrow C \neq \varnothing \wedge C \subseteq x.getCorpus$$

$$x.getState(C) = IMPOSSIBLE \leftarrow C \cap x.getCorpus = \varnothing$$

$x.getState(C) = POSSIBLE \leftarrow$

$x.getState(C) \neq IMPOSSIBLE \wedge x.getState(C) \neq KNOWN$

OPTIMISATION : On ne calculera *getCorpus* qu'une fois et on procèdera par élimination en parcourant *C* de conserve avec *getCorpus* pour y trouver un contre-exemple à *KNOWN* (un élément de *C* qui n'est pas dans *getCorpus*) et à *IMPOSSIBLE* (un élément de *C* qui est dans *getCorpus*).

Définition : $\_.getFilter(\_) : Descriptor \times DocumentObject^n \rightarrow (Descriptor \times State)^n$
Fonction de filtrage descendant récursivement dans les descripteurs connus et s'arrêtant aux descripteurs possibles et impossibles.

$x.getFilter(C) = \{(x, IMPOSSIBLE)\} \leftarrow x.getState(C) = IMPOSSIBLE$

$x.getFilter(C) = \{(x, POSSIBLE)\} \leftarrow x.getState(C) = POSSIBLE$

$x.getFilter(C) = \{(x, KNOWN)\} \cup \{(z,s) \mid specialization(x,y) \wedge (z,s) \in y.getFilter(C)\}$
$\leftarrow x.getState(C) = KNOWN$

Définition : $\_.getDescriptionContext(\_) : Facet^n \times Descriptor^{n \times n} \rightarrow (Descriptor \times State)^{n \times n}$
Obtention du contexte de description pour un ensemble de facettes et de sélections associées. On calcule d'abord le corpus global par intersection des corpus de chaque facette. Ensuite, on applique à chaque facette le filtre correpondant au corpus global.

$\{f_0,..f_n\}.getDescriptionContext(\{S_0,..S_n\}) = \{f_0.getFilter(C), .. f_n.getFilter(C)\}$
$\leftarrow C = f_0.getCorpus(S_0) \cap .. f_n.getCorpus(S_n)$

Optimisation 1 : L'utilisateur ne pouvant modifier qu'une sélection à la fois, il n'est pas nécessaire de recalculer les corpus des autres facettes. On conservera donc un « cache »

pour chaque facette. Celui-ci ne sera remis à jour qu'à la suite d'une modification de sélection.

Optimisation 2 : Chaque facette pouvant se situer sur un serveur différent, les appels aux méthodes *getFilter* se feront de manière parallèle.

Les figures 7.11 et 7.12 illustrent l'obtention du contexte de description et ses optimisations.

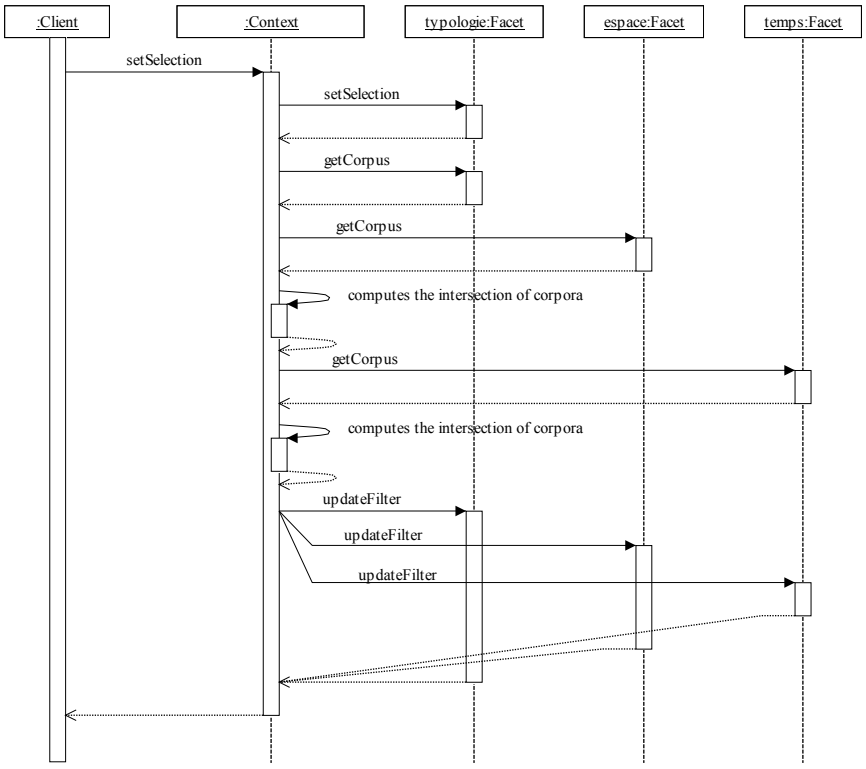


Figure 7.11 : Mise à jour des filtres dans les facettes « typologie », « espace » et « temps » après changement de sélection dans la facette typologie (diagramme de séquence UML).

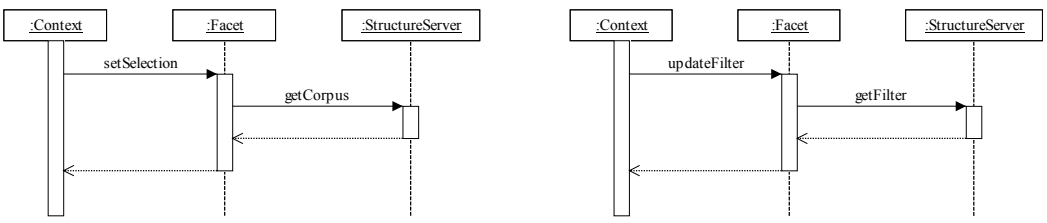


Figure 7.12 : Requêtes envoyées au serveur lors de la mise à jour des filtres (cf. Figure précédente).

c. Scénario de recherche de documents

Dans cette section, nous suivrons pas à pas un scénario d'interactions homme-machine mobilisées lors d'une recherche de document. Les schémas présenteront le réseau de description de la Figure 7.13 tel que filtré par le système *Porphyre* au cours d'une navigation.

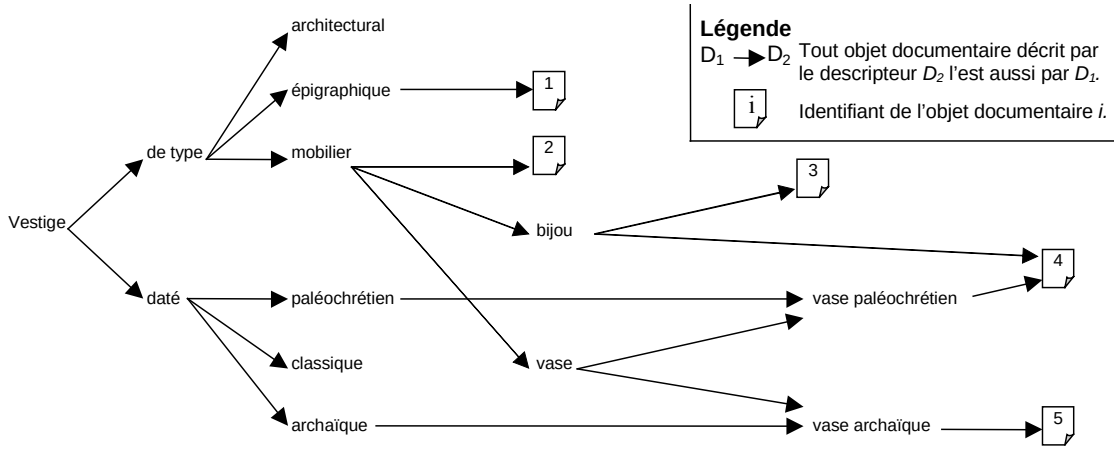
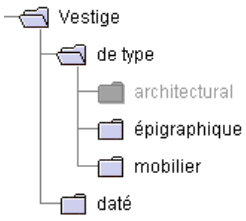


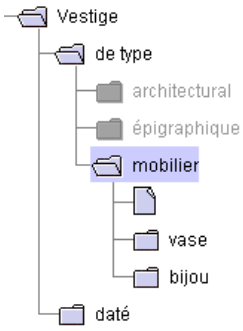
Figure 7.13 : Exemple de réseau de description

Étape 1 : Le corpus global traite de « vestige typé ». Les corpus plus spécialisés traitent de « vestige daté », de « vestige de type épigraphique » ou de « vestige de type mobilier » mais pas de « vestige de type architectural » (cette description ne correspond en effet à aucun document du corpus considéré). Après sélection par l'utilisateur de « mobilier », le système passe à l'étape 2.

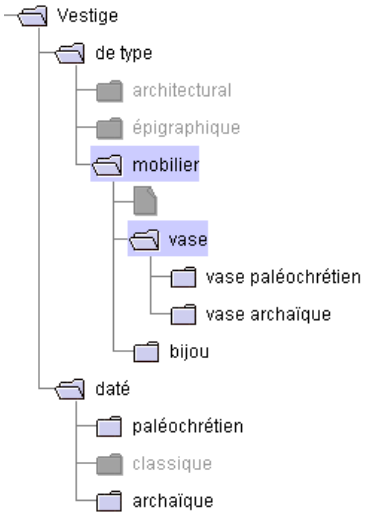


Étape 2 : Le corpus sélectionné traite de « vestige de type mobilier ». Cette description correspond exactement à celle de l'un des documents. Celui-ci est affiché. Des corpus plus spécialisés traitent de « bijou » ou de « vase ». Aucun ne traite de « vestige

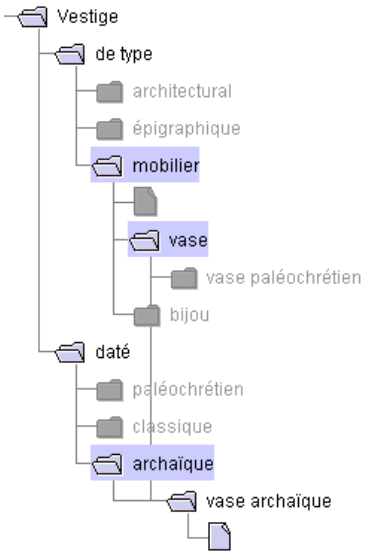
de type architectural » ou « épigraphique ». Après sélection par l'utilisateur de « vase », le système passe à l'étape 3.



Étape 3 : Le corpus sélectionné traite de « vestige de type vase (mobilier) » et de « vestige daté ». On remarque que le fait qu'il soit « daté » est inféré par l'ordinateur (car tous les documents du corpus traitant de « vase » traitent de « vestige daté »). Les corpus plus spécialisés traitent de « vase paléochrétien », de « vase archaïque », de « paléochrétien » ou « d'archaïque ». Aucun ne traite de vestiges de type « architectural », « épigraphique », « classique ». Le document qui était possible à l'étape précédente ne l'est plus. Après sélection par l'utilisateur « d'archaïque », le système passe à l'étape 4.



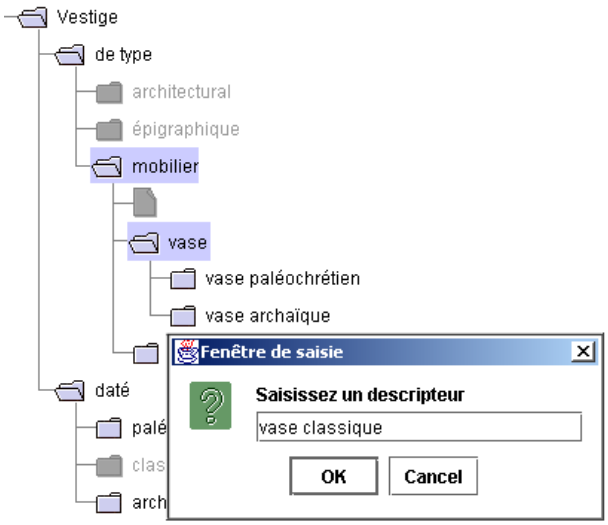
Étape 4 : Le corpus sélectionné traite de « vestige de type vase (mobilier) daté de l'époque archaïque ». Ce corpus ne contient qu'un seul document. Celui-ci est affiché.



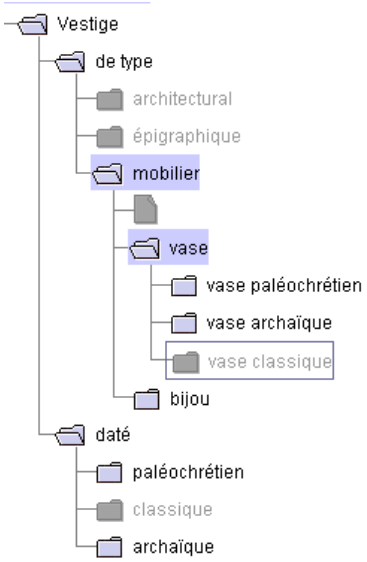
d. Scénario d'indexation de documents

Dans ce scénario, nous allons indexer un nouveau document traitant de « vase classique ». Tout d'abord, l'utilisateur doit trouver pour le nouveau document un corpus qui lui corresponde (mais pas ses sous-corpus). Comme le document traite de « mobilier »et de « vase », l'utilisateur peut suivre les étapes 1 et 2 (de la section précédente). Nous nous retrouvons alors à l'étape 3'.

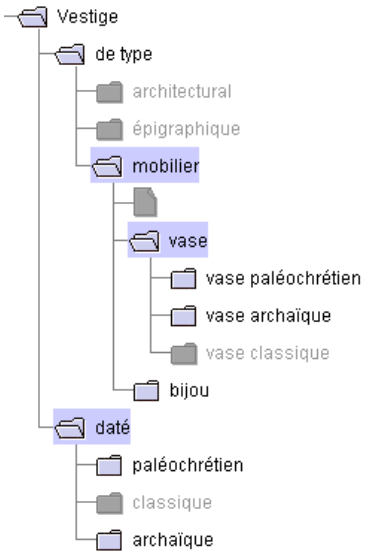
Étape 3' : Aucun des sous-corpus proposés (« vase paléochrétien », « vase archaïque », « paléochrétien » et « archaïque ») ne correspond au document à indexer. L'utilisateur crée donc un nouveau descripteur « vase classique » comme spécialisation de « vase », le système passe alors à l'étape 4'.



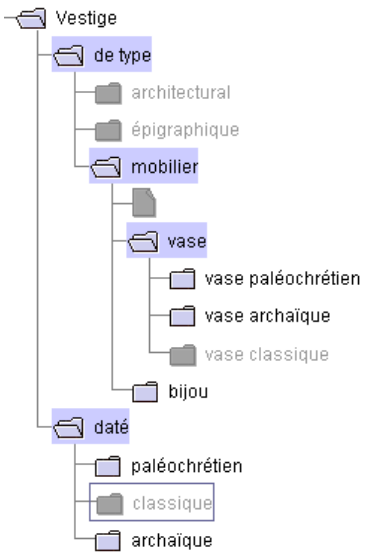
Étape 4' : Le nouveau descripteur « vase classique » apparaît comme n'étant affecté à aucun document. C'est alors que l'utilisateur crée l'identificateur de document « 6 » comme spécialisation de « vase classique ». Le système passe à l'étape 5'.



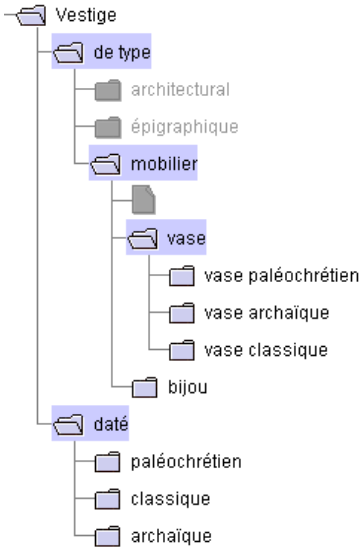
Étape 5' : Par conséquent, le système montre que « vase classique » décrit un corpus et que « daté » ne décrit plus l'ensemble des documents décrits par « vase ». Afin de comprendre cette incohérence potentielle, l'utilisateur sélectionne « daté ». Le système passe à l'étape 6'.



Étape 6' : Il apparaît que « vase classique » n'est pas déclaré comme « daté » et que « classique » n'est affecté à aucun document. L'utilisateur met alors le modèle à jour en reliant « vase classique » à « classique ». Le système passe à l'étape 7'.



Étape 7' : L'utilisateur a terminé d'indexer le nouveau document et « tout est pour le mieux dans le meilleur des mondes possibles »...



4. Bilan

Nous avons tout d'abord rappelé que l'organisation des bibliothèques (comme celle des documents) est en général basée sur un modèle arborescent, probablement en raison de l'influence de la philosophie d'Aristote sur notre manière de voir la connaissance. De manière à sortir de ce schéma trop simpliste, nous avons étudié certaines alternatives. Cependant, du fait qu'elles s'appuient toujours sur un modèle du domaine considéré comme fixe et extérieur, ces alternatives nous ont semblé telles quelles inapplicables à notre approche : celle d'une modélisation dynamique effectuée par les experts eux-mêmes et non par des tiers.

Nous avons ensuite proposé un modèle appelé « réseau de description » permettant à chaque expert de superposer au corpus sa propre structure, son propre point de vue. Une définition a alors été donnée de manière d'abord informelle puis à l'aide de spécifications algébriques.

Enfin nous avons vu, de manière formelle, puis sur des scenarii, comment on pouvait instrumenter l'expert dans sa tâche d'arpentage de l'espace documentaire grâce à un mécanisme de filtre.

Chapitre 8. Parcours de lecture

When the user is building a trail, he names it, inserts the name in his code book, and taps it out on his keyboard. Before him are the two items to be joined, projected onto adjacent viewing positions. […]Thereafter, at any time, when one of these items is in view, the other can be instantly recalled […]. Moreover, when numerous items have been thus joined together to form a trail, they can be reviewed in turn, rapidly or slowly, by deflecting a lever like that used for turning the pages of a book. It is exactly as though the physical items had been gathered together from widely separated sources and bound together to form a new book. It is more than this, for any item can be joined into numerous trails.
[…] Tapping a few keys projects the head of the trail. A lever runs through it at will, stopping at interesting items, going off on side excursions.
Vannevar Bush, As we may think, 7.

Comme nous l'avons vu dans le chapitre précédent, les réseaux de description permettent une navigation de corpus en sous-corpus (et inversement). Si cette navigation permet d'atteindre l'ensemble des corpus décrits, c'est souvent par une ennuyeuse série de « déselections » et de sélections de descripteurs. Ne serait-il pas utile d'offrir des « raccourcis » entre corpus, transversalement aux relations d'inclusion ? C'est pour répondre à cette attente, que nous proposons les *parcours de lecture*.

En tant que *trace,* nos parcours de lecture pourront être définis à la fois par les auteurs et les lecteurs des contenus documentaires. Dans le premier cas, ils pourront représenter, par exemple, la séquence des pages ou des illustrations. Dans le deuxième cas, ils s'apparenteront à un historique de lecture lié à une tâche donnée. Ils pourront dans les deux cas être édités, stockés, publiés et réutilisés ultérieurement.

Au cours de ce chapitre, nous présenterons, tout d'abord de manière informelle, nos choix de modélisation ainsi que les raisons qui les ont dictées. Dans un second temps, nous en donnerons une spécification algébrique. Ensuite, nous donnerons un aperçu des possibilités offertes par le modèle en déroulant un petit scénario d'utilisation. Enfin, nous montrerons en quoi notre modèle se distingue des travaux apparentés.

1. Principe

Nous définirons un parcours de lecture comme un *historique* parmi des *étapes de lecture*, permettant une *navigation.* Précisons maitenant chacun de ces aspects.

a. Historique

Le premier choix que nous devons faire porte sur l'*historique*. Dans le domaine de l'hypermédia, on distingue en général trois modèles [BieberEtWan94] : le modèle de la *pile* (utilisé dans les clients Web pour le « retour arrière »), le modèle *chronologique* (utilisé dans les serveurs Web comme « log »[74]) et le modèle de la *visite guidée*. Pour comprendre les différences entre les trois types d'historique, nous suivrons un exemple pas à pas (cf. Figure 8.1).

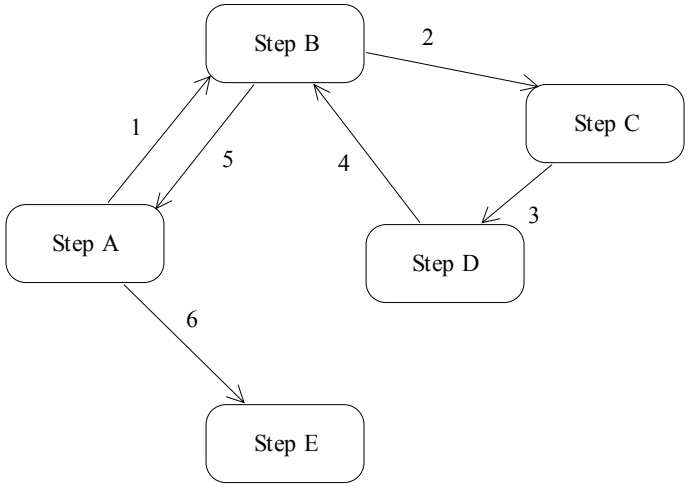


Figure 8.1 : Exemple de navigation entre des étapes de lecture (Diagramme d'état UML).

Supposons qu'un utilisateur passe par les étapes de lecture *A*, *B, C* puis *D*. Dans les trois modèles, l'historique sera [*A*, *B*, *C*, *D*].

Lorsque l'utilisateur retournera à l'étape *B*, le modèle de la pile indiquera [*A*, *B*], le modèle chronologique, [*A*, *B*, *C*, *D*, *B*] et le modèle de la visite guidée, [*A*, *B*, *C*, *D*].

[74] En supposant que le client Web n'ait pas de « cache ».

En retournant à l'étape *A*, l'utilisateur changera le premier historique en [*A*], le second en [*A*, *B*, *C*, *D*, *B*, *A*], et le troisième en [*A*, *B*, *C*, *D*].

Enfin en se rendant à l'étape *E*, l'utilisateur obtiendra pour historique :

- [*A*, *E*] dans le modèle de la pile,
- [*A, B, C, D, B, A, E*] dans le modèle chronologique,
- et [*A, B, C, D, E*] dans le modèle de la visite guidée.

Le modèle de la visite guidée, en stockant l'ensemble des étapes de lecture dans l'ordre de leur première visite et en négligeant les retours en arrière, nous semble préférable du point de vue de la charge cognitive et de la charge computationnelle.

b. Etape de lecture

Ayant défini pour les parcours de lecture ce que nous appelions un historique, nous avons maintenant à préciser ce que nous appellerons une *étape de lecture*.

On serait sans doute tenté d'assimiler une étape de lecture à un objet documentaire (source, fragment, note). Cependant une telle définition nous priverait du contexte documentaire de lecture[75] – l'un des aspects les plus intéressants des réseaux de description.

Pour autant, prendre comme étape le contexte documentaire de lecture serait assez peu judicieux. En effet, si l'on considère les opérations définies pour la navigation dans un réseau de description comme autant de systèmes, ce contexte serait une « sortie » mais jamais une « entrée ». Dit autrement, il serait possible de poursuivre une navigation dans les réseaux de description par une navigation dans les parcours de lecture mais pas l'inverse.

Tenant compte de l'objection précédente, on serait amené à définir une étape de lecture comme un corpus de documents. Cependant, rechercher une étape parmi les *N* étapes disponibles reviendrait à effectuer *N* comparaisons d'ensembles ! En outre, ne considérer que le corpus, néglige le fait qu'à un instant donné l'utilisateur concentre son

[75] Ensemble des objets documentaires affichés à un instant donné par le client de *Porphyre*.

attention sur un seul des objets documentaires présents à l'écran (ce qui se traduit par une activation de la fenêtre correspondante).

Les remarques précédentes, nous conduisent à définir une étape dans un parcours de lecture comme un objet documentaire parmi un corpus. Ainsi, rechercher une étape revient à comparer des « localisations d'objets documentaires[76] » entre elles. Une fois l'étape trouvée, on peut récupérer le corpus dans lequel l'élément documentaire doit être consulté.

Reste à préciser comment le corpus sera désigné : en intension (par ce que l'on a appelé une sélection) ou en extension (par la liste des objets documentaires le composant). Si la première est beaucoup plus concise, elle présente un inconvénient majeur, celui d'avoir un résultat dynamique. Après modification du réseau de description, le corpus obtenu pourrait même ne plus contenir l'élément documentaire cherché ! Les corpus seront donc notés en extension.

c. Navigation

Dans le modèle ainsi défini, plusieurs parcours de lecture pourront se croiser en un même objet documentaire. Par conséquent, le système devra indiquer pour l'objet documentaire activé la liste de ses parcours (pour les facettes connectées).

Après sélection de l'un de ces parcours, étant donné qu'un objet documentaire ne peut apparaître qu'une fois dans un parcours de lecture, le lecteur pourra choisir sans ambiguïté l'étape précédente, l'étape suivante ou l'origine du parcours.

2. Spécification

La Figure 8.2 nous permet de préciser notre modèle. A l'intèrieur d'une facette, il sera possible de définir des *Parcours*. Un parcours correspondra à la séquence de plusieurs objets documentaires. Inversement un objet documentaire pourra apparaître dans plusieurs parcours. Les objets documentaires référencés par les parcours pouvant

[76] Cf. Chapitre 6.

être distants[77], il n'y aura aucune contrainte d'intégrité référentielle sur eux. On appellera « *Etape* » le couple unissant un parcours et un objet documentaire. Chaque étape correspondant à contexte de lecture, elle comportera un *corpus* d'objets documentaires (pouvant eux aussi être distants).

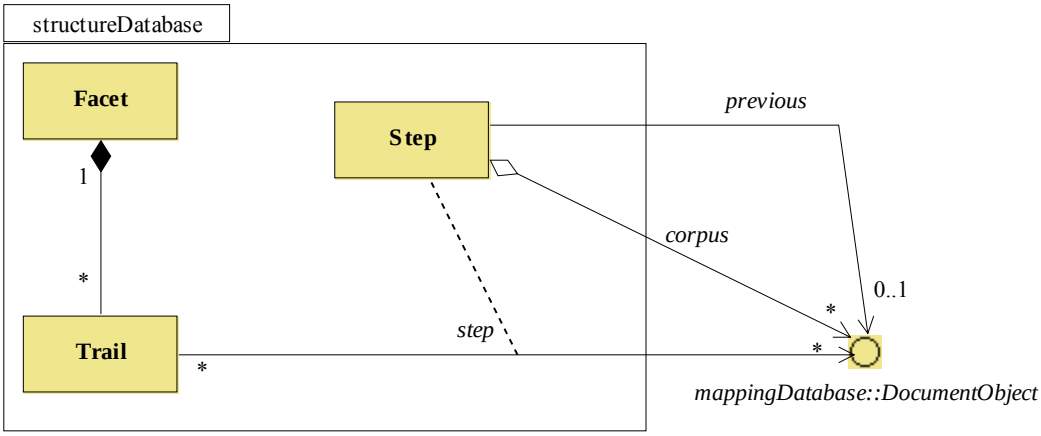


Figure 8.2 : Parcours de lecture (diagramme de classe UML)

Plus formellement, nous aurons affaire aux ensembles suivants : *Facet, Trail, Step, DocumentObject*. De même que dans le chapitre précédent, nous utiliserons des spécifications algébriques[78] pour préciser :

- le schéma des données à stocker (primitives),
- les contraintes supplémentaires que ces données doivent respecter (contraintes),
- les requêtes complexes qui seront effectués sur ces données (définitions).

Primitive : *step*(_,_) : *Trail* × *DocumentObject* → *Boolean*

[77] Grâce à l'utilisation des « localisations d'objets documentaires (cf. Chapitre 6).

[78] Voir en annexe l'explication de la syntaxe utilisée.

Axiome :

On appellera *Step* l'ensemble des couples (*t,d*) pour laquelle la relation homonyme sera vraie.

$Step = \{(t,d) | \ step(t,d)\}$

Primitive : *_.getPrevious* : $Step \rightarrow DocumentObject \cup \{NULL\}$

Le fait qu'il s'agisse d'une fonction et non d'une relation traduit la contrainte d'intégrité fonctionnelle suivante : pour une étape donnée, il existe au plus un *DocumentObject* précédent.

Contrainte : Unicité du prochain

Deux étapes d'un même parcours ne peuvent avoir le même prédécesseur.

$\perp \leftarrow (t,d_1).getPrevious = d_0 \wedge (t,d_2).getPrevious = d_0 \ \wedge d_1 \neq d_2$

Contrainte : Intégrité référentielle dans le parcours

Si une étape a un prédécesseur, celui-ci est forcément issu du même parcours.

$\perp \leftarrow (t,d_1).getPrevious = d_0 \wedge \neg \ step(t,d_0)$

Définition : *_.getHome* : $Trail \rightarrow \ DocumentObject$

On appelle *origine* d'un parcours, l'étape de ce parcours ayant pour prédécesseur *NULL*.

$t.getHome = d \leftarrow (t,d).getPrevious = NULL$

Définition : $\_.getNext : Step \rightarrow DocumentObject \cup \{NULL\}$

La fonction de succession est l'inverse de celle de précédence. Dans le cas, où il n'existe aucune étape de t ayant d_0 pour predecesseur, on dira que le successeur de l'étape (t,d_0) est *NULL*.

$(t,d_0).getNext = d_1 \leftarrow (t,d_1).getPrevious = d_0$

$(t,d_0).getNext = NULL \leftarrow \neg ((t,\_).getPrevious = d_0)$

Primitive : $\_.getFacet : Trail \rightarrow Facet$

Le fait qu'il s'agisse d'une fonction traduit la contrainte d'intégrité fonctionnelle suivante : un parcours de lecture appartient à une seule facette.

Définition : $\_.getTrails(\_) : Facet \times DocumentObject \rightarrow Trail^n$

Pour une facette donnée, permet d'obtenir tous les parcours ayant une étape passant par un objet documentaire donné.

$f.getTrails(d) = \{t |\ t.getFacet = f \wedge (t,d).getPrevious = \_ \}$

Primitive : $\_.getCorpus : Step \rightarrow DocumentObject^n$

Cette fonction permet d'obtenir le corpus correspondant à une étape de lecture. On peut alors appliquer le *getFilter* défini dans les réseaux de description. Ainsi, les deux types de navigation peuvent s'enchaîner.

3. Scénario : Feuilleter un ouvrage

Dans l'exemple de la figure 8.3, nous disposons de deux parcours de lecture pour feuilleter un ouvrage (le n°12) : le premier donnant l'enchaînement des paragraphes et le second celui des figures. Un troisième parcours de lecture correspond à l'enchaînement des références bibliographiques d'un autre ouvrage (le n°30) citant le premier. Dans le premier parcours, chaque paragraphe est lu en contexte avec les figures qui y sont référencées. Dans le second, c'est l'inverse. Dans le troisième, chaque passage référencé est lu en contexte avec ceux qui s'y réfèrent.

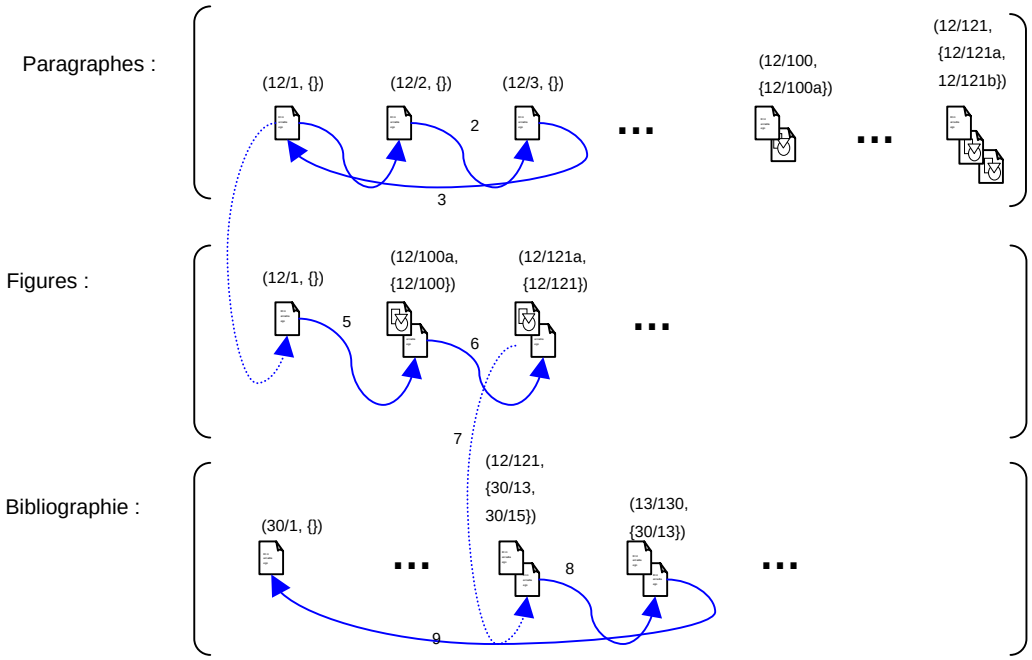


Figure 8.3 : Exemple de parcours de lecture

Supposons que le lecteur choisisse de parcourir l'ouvrage n°12, il se positionne sur le premier objet documentaire (12/1). Il apprend qu'il existe deux parcours documentaires qui y passent : un premier nommé « paragraphes » et un second « figures ». Il choisit paragraphe et passe à l'étape suivante.

A la lecture du nouveau paragraphe (12/2) et de son successeur (12/3) dans le parcours, le lecteur constate que l'ouvrage est susceptible de l'intéresser. Il décide de feuilleter les figures de l'ouvrage. Pour ce faire, il revient à l'origine du parcours « paragraphes » et choisit le parcours « figures ».

En se rendant à l'étape suivante, le lecteur découvre la figure 12/100a, puis 12/121a, toutes deux accompagnées de leur paragraphe explicatif. Intéressé par la figure 12/121a, il sélectionne le paragraphe 12/121 pour le lire.

Il apprend alors qu'un autre parcours, nommé « bibliographie », passe par ce paragraphe. Intéressé par les commentaires 30/13 et 30/15 qui sont faits du 12/121, il continue le parcours. Un autre ouvrage (n°13) très pertinent y est référencé et commenté. Intrigué par la bibliographie de l'ouvrage n°30, le lecteur décide de lire ce dernier *in extenso*, et va donc à l'origine du parcours de lecture.

Notons qu'en feuilletant ainsi le corpus, notre lecteur a tracé un nouveau parcours. S'il le souhaite, il pourra lui donner un nom et le conserver pour un usage ultérieur.

4. Originalité du modèle

Nombreux furent les systèmes hypermédia à proposer une implémentation de la notion de parcours introduite par Vannever Bush [Bush45] : une séquence d'objets documentaires créée par le lecteur, nommée, sauvegardée et publiée.

Certains, conscients du danger de désorientation que représentait une lecture « objet documentaire » par « objet documentaire », proposèrent comme étape de lecture non pas un objet documentaire isolé mais un ensemble de documents [TriggEtWeiser86, Trigg88, Maurer96]. Cependant, leurs parcours ne pouvaient se croiser que s'ils utilisaient explicitement la même référence[79] pour désigner cet ensemble.

Au contraire, la notion de parcours de lecture nous semblant fortement liée à celle du point de vue, il nous semblait indispensable de permettre le croisement de parcours issus de deux points de vue différents (donc ne dépendant l'un de l'autre que par l'intermédiaire du corpus). Le modèle proposé par nos soins permet d'y parvenir tout en gardant une complexité algorithmique très raisonnable.

[79] Cette référence est appelée, suivant les modèles, « noeud table-des-matières » [TriggEtWeiser86], « dessus-de-table » [Trigg88] ou « grappe » [Maurer96].

3ème partie : Études de cas

Chapitre 9. La *Chronique des fouilles* du BCH*

L'École française d'Athènes publie une douzaine de séries de monographies ainsi qu'une revue : le *Bulletin de correspondance hellénique* [BCH]. Cette revue annuelle est consacrée à des articles de synthèse, à des informations sur les activités de l'École et à la publication du matériel archéologique. Une de ses sections, la *Chronique des fouilles* a pour mission particulière de signaler aux lecteurs toutes les « nouveautés » archéologiques[80] sur lesquelles des informations fiables ont été obtenues au cours de l'année.

L'Ecole française d'Athènes, souhaitant une valorisation accrue de ce corpus assez unique en son genre, projeta en 1996 sa mise en ligne[81]. Lors d'une étude de faisabilité, menée par nos soins[82] en 1998, il apparut que la richesse de sa mise en forme nécessitait un nouveau type d'outil informatique. Cet outil fut ensuite prototypé au cours de notre DEA puis de notre thèse pour devenir le système *Porphyre.* Ce que nous présentons dans ce chapitre concerne l'expérimentation menée à l'automne 2000 de gestion de la *Chronique des fouilles* à l'aide de *Porphyre* (expérimentation qui fut suivie d'une démonstration publique lors de la table-ronde interdisciplinaire de novembre 2000 à Athènes).

Dans un premier temps, nous procèderons à une étude de l'existant, en analysant en particulier les caractéristiques structurelles du corpus et leur évolution dans le temps. Dans un deuxième temps, nous détaillerons l'expérimentation proprement dite. Enfin, dans un troisième temps, nous verrons quels furent les retours de l'expérience, tant au niveau du prototype que de la manière de traiter le cas de la chronique.

---

* Certains éléments de ce chapitre ont été présentés lors de la table-ronde "Sémantique et Archéologie" [Benel00] et de la journée d'étude sur les bibliothèques numériques organisée par l'Ecole française d'Athènes, la Maison de l'Orient Méditerranéen et l'Université Lyon II [Benel02b].

[80] Fouilles, prospections, trouvailles fortuites, restaurations, muséologie, publications de matériel inédit.

[81] Etude inscrite au plan quadriennal 1996-1999.

[82] Lors d'un stage d'élève ingénieur en 4ème année INSA.

1. Etude de l'existant

a. Présentation du corpus

La chronique fait son apparition dans le bulletin de 1920. Appelée « Chronique des fouilles dans l'Orient hellénique », elle a alors pour portée géographique l'ensemble de la Grèce antique. En 1936, elle adopte son titre actuel de « Chronique des fouilles en Grèce » et se cantonne depuis aux frontières actuelles de la Grèce. En 1959, parallèlement à la chronique en Grèce, apparaît la « Chronique des fouilles à Chypre » puis en 2000, la « Chronique des fouilles dans le Bosphore cimmérien ». Couvrant indifféremment, à l'origine, les fouilles de l'École et les autres, la chronique se scinde, en 1940, en deux rubriques traitant respectivement des unes et des autres. Enfin, à partir de 1970, les travaux de l'École n'apparaissent plus dans la chronique que sous forme de références à une section autonome du bulletin créée pour l'occasion.

Par son histoire et son volume d'informations, la Chronique des fouilles représente un outil pratiquement unique pour les chercheurs : quatre-vingt ans d'activité archéologique en Grèce et trente ans à Chypre. De plus, malgré cette quasi-exhaustivité, le volume du corpus (cf. Tableau 9.1) est suffisamment limité pour envisager sa numérisation intégrale (textes re-saisis, figures numérisées d'après originaux) dans des délais et des budgets raisonnables.

NOMBRE DE PAGES	OCCUPATION MOYENNE DES PAGES PAR LES FIGURES	NOMBRE DE CARACTÈRES
12.000 ± 500	(55,0 ± 2,5) %	30.000.000 ± 500.000

Tableau 9.1 : Volumétrie de la *Chronique des fouilles*[83]

[83] Les volumes indiqués ici s'appuient sur l'inventaire réalisé par nos soins en mai 1998 sur les 108 chroniques de fouilles disponibles, à savoir les 70 numéros de la Chronique des fouilles en Grèce (et en Orient Hellénique) de 1920 à 1995 (sachant qu'il n'en existe pas pour 1932 et 1946, et que sont regroupés respectivement ceux de 1940 et 1941, 1942 et 1943, 1947 et 1948) et les 38 numéros de la Chronique des fouilles à Chypre de 1959 à 1996.

b. « Chronique » ou « Topique » des fouilles ?

La *Chronique* est un corpus constitué de telle manière que chaque passage soit compréhensible indépendamment des autres (chacun traitant en général des découvertes annuelles pour un site géographique donné). Cependant, pour offrir une compréhension plus approfondie, des liens entre ces passages sont assurés par une structure complexe.

Cette structure est avant tout temporelle (comme le nom de *Chronique*[84] l'indique) : chaque livraison correspondant à une année de fouilles. Ensuite, dans chacune de ces livraisons, une hiérarchie des titres particulièrement profonde (au moins sur cinq niveaux) permet de regrouper les passages essentiellement par rapport à la géographie des sites dont ils traitent (donnant une impression de « zoom » d'un niveau à l'autre).

Il est intéressant de remarquer que cette deuxième structure est en partie invariante. Ainsi, il est envisageable à partir du corpus de la *Chronique des fouilles* d'obtenir automatiquement un nouveau corpus que l'on pourrait nommer « *Topique*[85] *des fouilles* » (cf. Figure 9.1). Ce traitement viendrait d'ailleurs généraliser une pratique existante consistant à rassembler les « tirés à part » concernant un site.

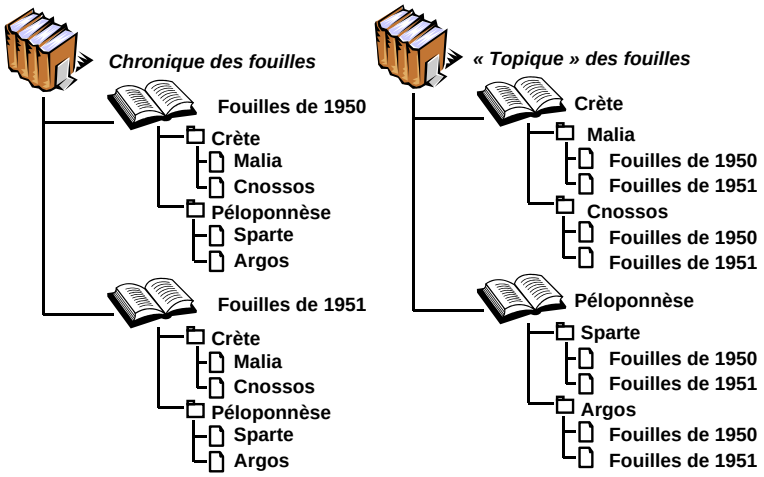


Figure 9.1 : La *« Topique des fouilles»,* virtuellement présente dans la *Chronique*.

[84] En Grec, *Chronos* signifie « le temps ».

[85] En Grec, *Topos* signifie « le lieu ».

Afin de modéliser cette structuration multiple du corpus, il faudrait « décorréler » la dimension spatiale et temporelle en deux facettes (cf. Figure 9.2). La recomposition en « chronique », « topique » ou toute autre combinaison des deux dimensions pourrait ainsi être calculée à la demande.

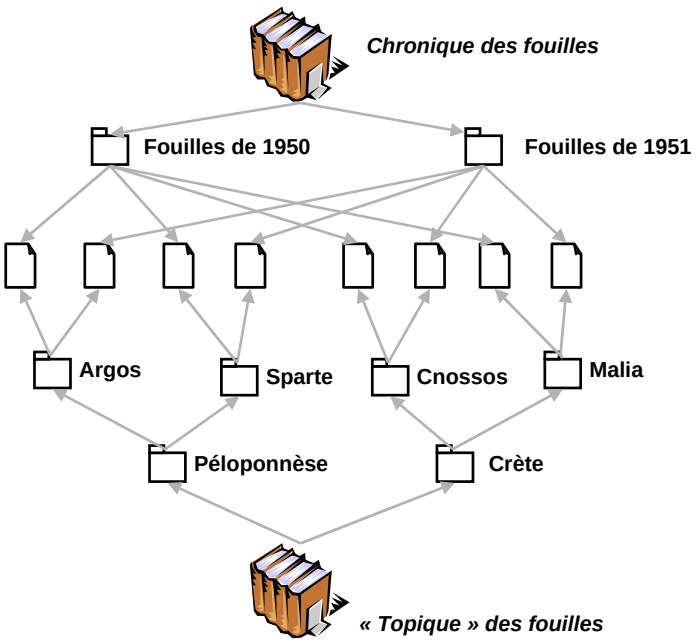


Figure 9.2 : Réseau de description présentant deux facettes pour un même corpus

2. Expérimentation

Notre expérimentation, menée en automne 2000, portait sur les quelques chroniques disponibles en texte intégral[86]. Nous basant alors sur la typologie courante distinguant dans le document numérique ses structures physique, logique et sémantique [NanardEtAl96], nous avions considéré :

- que la structure par page devait être omise (puisque en tant que structure physique, elle était calculable en fonction de la structure logique),
- que la structure actuelle en livraisons et en sections constituerait la structure logique, et qu'elle serait utilisée pour le stockage du document (en XML),

[86] Les chroniques postérieures à 1994 sont en effet archivées au format Micro$oft Word par le service des publications. Cependant, il s'agit uniquement des brouillons envoyés au maquettiste, c'est-à-dire avant insertion des figures, mise en page et corrections diverses.

- enfin, que la structure géographique et temporelle constituerait la structure sémantique.

La figure 9.3 correspond à la démonstration effectuée lors de la table ronde « Sémantique et Archéologie » (Athènes, novembre 2000). Elle montre que le prototype d'alors permettait une double navigation : à travers les documents XML et les réseaux de description. Elle présente également le réseau de description obtenu par le mode opératoire suivant :

1. convertir automatiquement la structure XML en réseau de description,
2. puis, dé-coreller à la main les dimensions temporelles et spatiales,
3. enfin, fusionner, toujours à la main, les structures spatiales de chaque livraison afin d'obtenir une sorte de thesaurus géographique contextualisé par des récits de fouille.

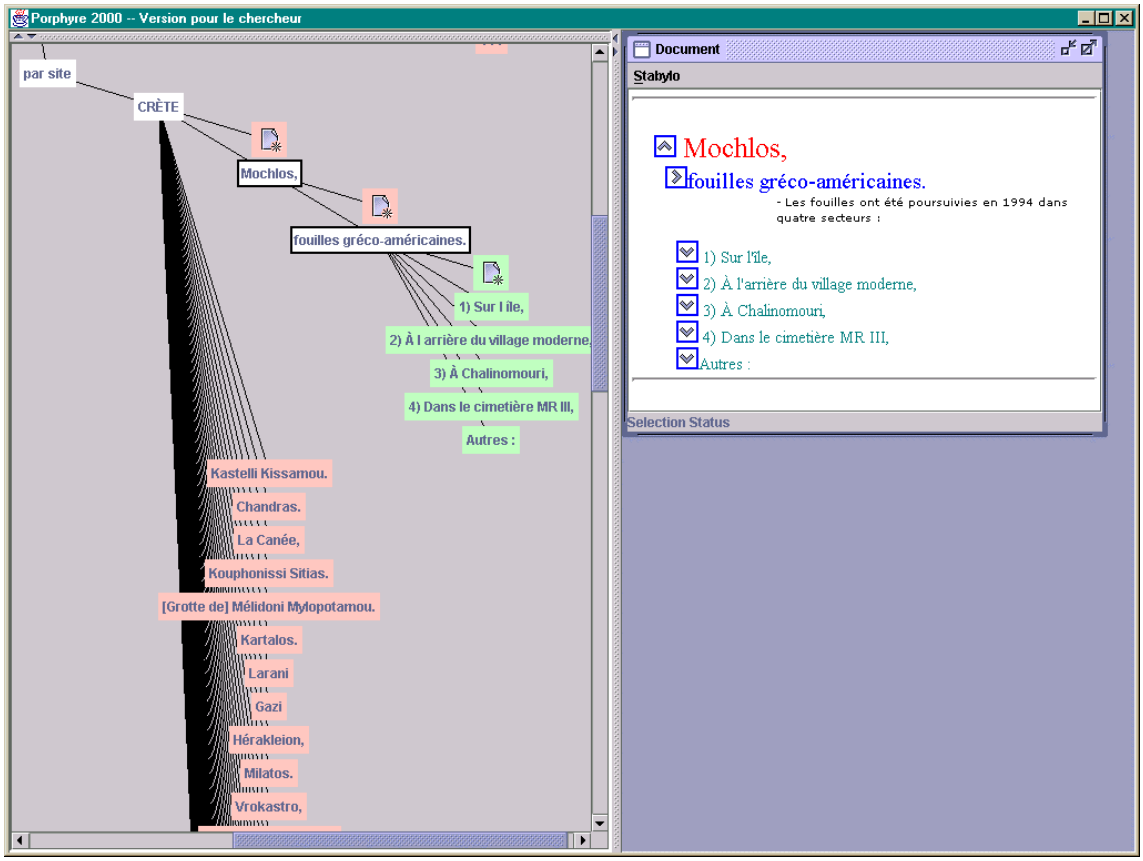


Figure 9.3 : Lecture avec *Porphyre 2000* d'un extrait de la *Chronique des fouilles*

3. Retour d'expérience

a. Limites rencontrées

Pour nous inscrire dans le cadre de la méthode poppérienne, nous ne nous appesantirons pas ici sur les résultats positifs de l'expérimentation et des démonstrations qui suivirent, mais chercherons plutôt les problèmes rencontrés.

Tout d'abord, il est apparu assez vite que la redondance des structures posait un problème cognitif. N'était-il pas « contre-intuitif » que le descripteur « Thrace » désigne dans la structure XML les fouilles d'une année, et dans le réseau de description celles de toutes les années ? Une autre erreur avait été de mépriser, comme on le fait souvent dans le domaine de la documentique, la notion de « page ». Que faire alors des références bibliographiques (depuis l'intérieur et l'extérieur du corpus), toutes de la forme « BCH 1995, p.409-410 » ? De plus, en raison de la différence de sujets traités par les différents passages, nous avions abandonné toute notion de séquence. Or les références croisées (« *ibid*. », « *op. cit.* ») ne peuvent être interprétées que si l'on peut parcourir les passages précédents. Enfin, il s'est avéré extrêmement éprouvant de fusionner les structures géographiques à la main.

b. Solutions proposées

Les évolutions du système *Porphyre* ont tenté de tenir compte des limites rencontrées. Tout d'abord, pour éviter la redondance entre la structure des fichiers XML et celle des réseaux de description, les fichiers XML ne sont plus, désormais, stockés tels quels dans *Porphyre,* mais convertis en réseau de description et objets documentaires (ce qui optimise d'ailleurs la plupart des traitements). Ensuite, pour introduire dans *Porphyre* la notion de séquence, nous avons mis en place les parcours de lecture (cf. Chapitre 8). Enfin, au sujet de la fusion semi-automatique de structures, nous disposons déjà d'un certain nombre de pistes (cf. Chapitre 12).

En ce qui concerne l'organisation du corpus, il sera souhaitable lors de la numérisation effective de la chronique de prendre pour objets documentaires des textes

courts, des photographies, des plans, et de les organiser à l'aide des structures suivantes (cf. Figure 9.4) :

- un parcours de lecture modélisant la séquence de ces objets documentaires,
- un réseau de description modélisant leur description bibliographique (Revue, livraison, page), et reprenant ainsi indirectement la dimension temporelle (année de fouille),
- enfin, un deuxième réseau obtenu par fusion des structures spatiales de chaque livraison.

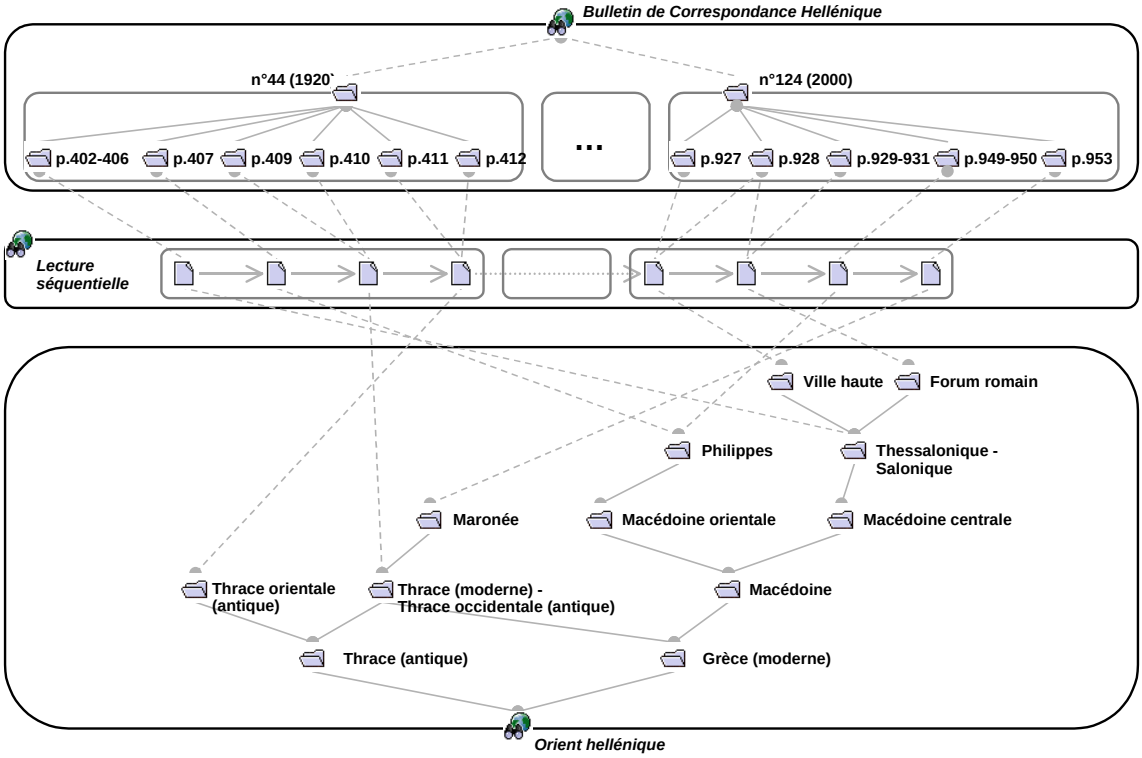


Figure 9.4 : Prise en compte du retour d'expérience : nouvelle structuration de la *Chronique des fouilles* à l'aide de Porphyre

En ce qui concerne les perspectives déjà évoquées d'assistance à la fusion, notons que l'exemple de la figure ci-dessus est particulièrement représentatif de la complexité de l'opération. En effet, nous avons dû ici tenir compte de changements de nom (Thessalonique/Salonique), de différences de granularité (ville/quartier) et de changements structurels encore plus radicaux (la Thrace moderne correspondant à la Thrace occidentale antique puisque la Thrace orientale antique est aujourd'hui en Turquie). On

conçoit ici aisément pourquoi il serait vain de proposer pour cette tâche une automatisation dépassant le cadre d'assistants interactifs.

Chapitre 10. CEFAEL : Collections de l'Ecole française d'Athènes en ligne*

Comme nous l'avons vu dans le chapitre précédent, l'École française d'Athènes publie une douzaine de collections (séries de monographies et revue). Depuis 1877, ce sont près de 570 volumes, soit 250.000 pages, qui ont été édités par l'Ecole. Si un tel corpus représente un intérêt indéniable pour l'historien des sciences, il n'en est pas moins important pour l'archéologue. En effet, la fouille archéologique présente la particularité de détruire les couches qu'elle étudie. Que reste-t-il alors de son objet d'étude ? Le carnet de fouille, l'article, la monographie… Ainsi, un article de 1877, malgré les révolutions théoriques et de méthodologiques qu'a pu connaître la discipline depuis, reste-t-il un substitut incontournable des vestiges qu'il décrit.

A l'heure où des fondations américaines mettent en place d'immenses bibliothèques numériques (comme JSTOR[87]) portant sur la rétrospective des revues en Sciences Humaines, le Ministère de la Recherche a souhaité encourager les expérimentations technologiques permettant à terme, au niveau français ou européen, des alternatives publiques. C'est ainsi qu'en décembre 2001 le projet de mise en ligne des collections de l'Ecole a reçu le soutien financier du « Plan de numérisation des publications en SHS ». Aujourd'hui, le portail CEFAEL[88] permet, à travers de multiples structures hypermédia, de feuilleter gratuitement sur la Toile l'intégralité des pages du corpus (sous forme de fac-similés).

La nature du projet nécessitait une valorisation immédiate du corpus à l'aide de technologies éprouvées. Cependant, comme nous allons le voir dans ce chapitre, il a été possible d'expérimenter la gestion avec *Porphyre* d'une partie de ce corpus. Dans une première partie, nous étudierons qu'elles sont les différentes structures hypermédia dont a besoin le lecteur. Dans une deuxième partie, nous présenterons la chaîne de numérisa-

---

* Des parties de ce chapitre ont fait l'objet d'une conférence lors de la journée d'étude sur les bibliothèques numériques [Benel02b].

87 http://www.jstor.org

88 http://cefael.efa.gr

tion et de diffusion mise en œuvre dans le projet CEFAEL. Ensuite, dans une troisième partie, nous exposerons le protocole expérimental proprement dit. Enfin dans une quatrième partie, nous verrons les problèmes rencontrés et les solutions proposées.

1. Etude des besoins

Afin de définir les différentes structures hypermédia nécessaires à l'interprétation des collections de l'Ecole, nous allons tenter d'identifier les « points de vue » de différents acteurs intervenant sur une même page du corpus (cf. Figure 10.1).

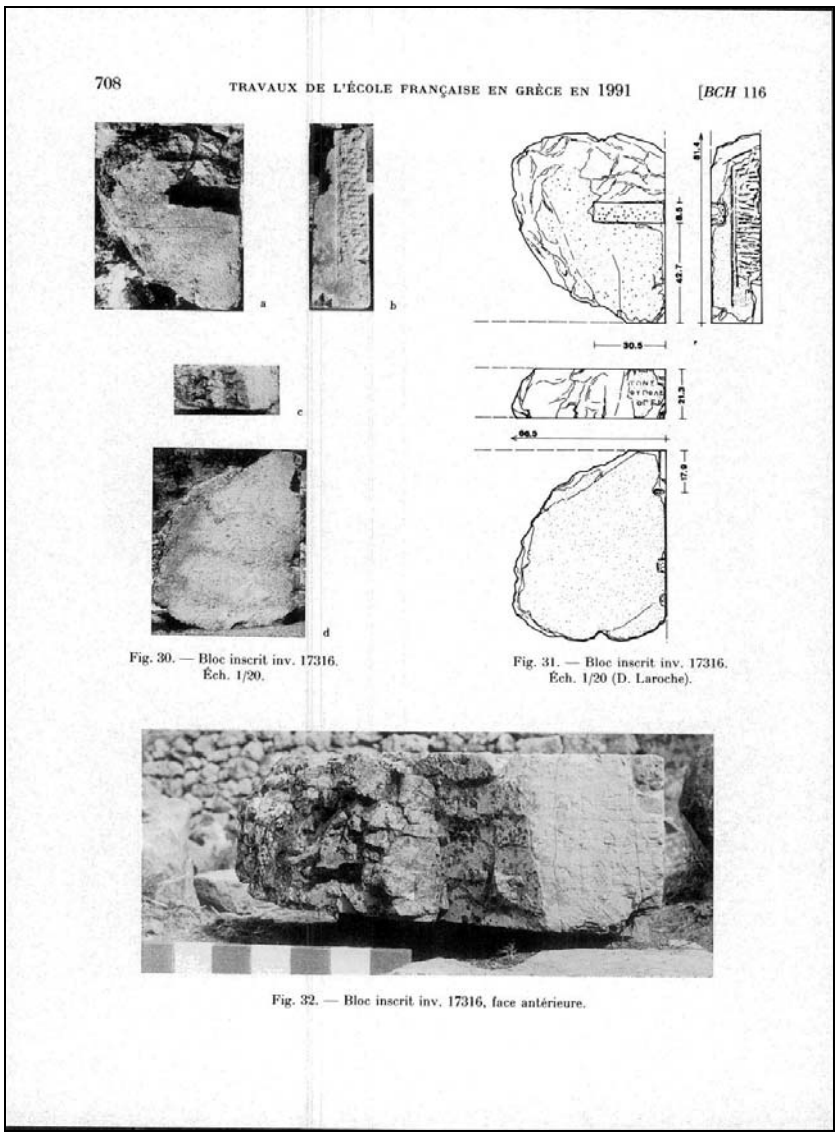

708
TRAVAUX DE L'ÉCOLE FRANÇAISE EN GRÈCE EN 1991
[BCH 116
a
b
c
d
30.5
Fig. 30. — Bloc inscrit inv. 17316. Éch. 1/20.
Fig. 31. — Bloc inscrit inv. 17316. Éch. 1/20 (D. Laroche).
Fig. 32. — Bloc inscrit inv. 17316, face antérieure.

Figure 10.1 : Une page à étudier sous différents points de vue.

a. Maquettiste

La première structuration du corpus est donnée par le maquettiste : il s'agit de la pagination. Chaque page est ainsi désignée sans ambiguïté par le triplet « Collection/Volume/Folio ». Cette nomenclature arborescente permet ainsi de nommer la page choisie « BCH/116/708 » (cf. Figure 10.2). On peut utiliser des folios spéciaux pour ceux habituellement en chiffre romain (pages préliminaires) et pour les pages non foliotées (pages finales, dépliants, planches…). Notons qu'il n'est pas indispensable d'introduire le niveau du tome puisque la pagination est continue d'un tome au suivant.

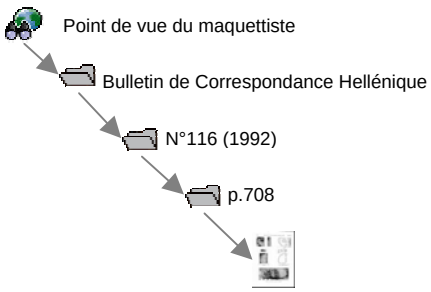


Figure 10.2 : Extrait de la facette du maquettiste (Réseau de description *Porphyre*)

b. Bibliothèque

Si la structure précédente suffit à référencer l'ensemble du corpus, le chercheur a cependant besoin d'autres structures pour y accéder. L'une de ces structures est celle qui apparaît dans le catalogue de la bibliothèque. Cette structure identifie au sein des volumes des éléments que l'on appellera « publications » (articles de recherche, rapports, chroniques…). Ces publications ont pour attribut une date et un ou plusieurs auteurs. Notre page d'exemple (cf. Figure 10.3) appartient à un rapport sur les travaux de l'Ecole à Delphes, daté de 1991, et cosigné par les huit auteurs indiqués.

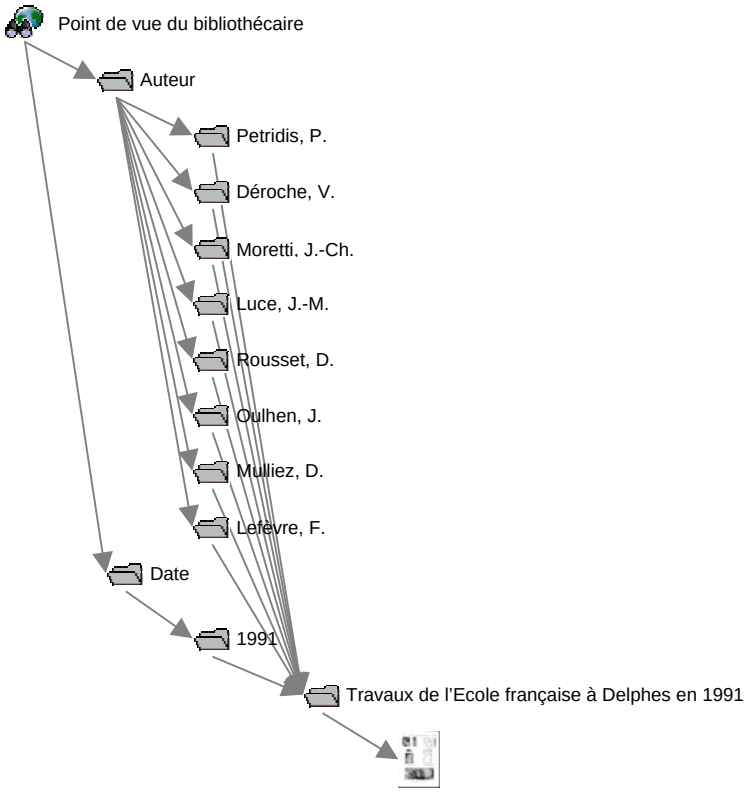


Figure 10.3 : Extrait de la facette du bibliothécaire (Réseau de description *Porphyre*)

c. Photothèque/Planothèque

L'Ecole dispose d'un fond de près de 500.000 photographies et plans datant de la fin du XIX siècle à nos jours. Ce fond comprenant entre autres les figures publiées dans les collections, on peut considérer que la structure du fond est aussi structure de la collection. Ainsi, la Figure 10.4 montre-t-elle que notre page d'exemple comprend deux figures correspondant aux photographies d'archive « R3879-007 » et « L9689-030 ». Chacune de ces photographies peut être décrite par un certain nombre de « méta-données », telles que leur auteur (Jean-Charles Moretti) et leur date de prise de vue (1991).

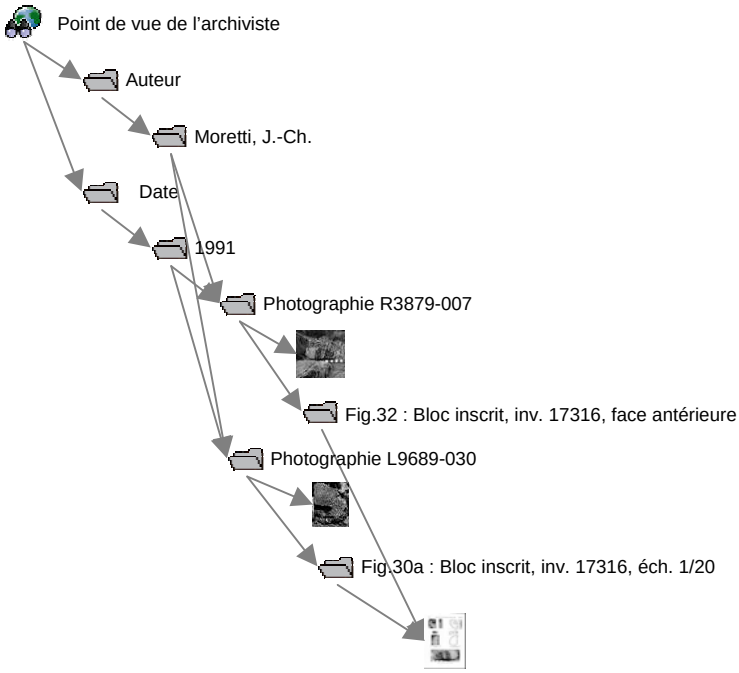


Figure 10.4 : Extrait de la facette de l'archiviste (Réseau de description *Porphyre*)

d. Equipe de fouille

Les trois premières structures étudiées sont loin d'être exhaustives. En effet le corpus est appelé à être structuré par chacun de ses lecteurs. Un exemple intéressant nous est donné par l'équipe de fouille de Roland Etienne. Cette équipe travaille actuellement à analyser la bibliographie concernant le sanctuaire de Délos en fonction de la position spatiale de chacun des vestiges décrits. La plupart de cette bibliographie étant contenue dans les collections de l'Ecole, on peut donc considérer que l'on est en présence d'une nouvelle structure du corpus. De la même manière, notre page d'exemple pourrait entrer dans une structure correspondant à la géographie du site de Delphes (cf. Figure 10.5).

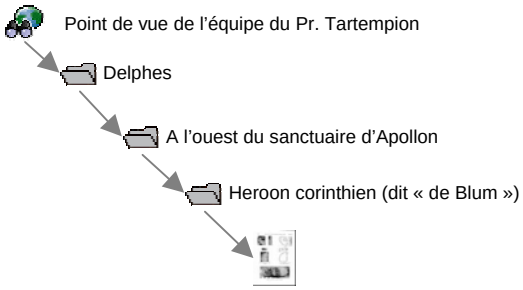


Figure 10.5 : Extrait de la facette d'une équipe de fouille (Réseau de description *Porphyre*)

2. Numérisation et valorisation

La numérisation du corpus démarra en mai 2001, à Lyon, sur le site de la plateforme technologique du CNRS à la Maison de l'Orient et de la Méditerranée [MOM]. Les corpus complets, disponibles en bibliothèques, ne pouvant être ravis aux lecteurs, il fut nécessaire, pour reconstituer les collections, de faire venir des volumes d'Athènes (EFA, éditeur), de Limoges (Bontemps, imprimeur) et de Paris (De Boccard, distributeur).

Chaque volume fut, préalablement à la numérisation, décrit dans une base de données (nombre de pages foliotées, nombre de planches, etc.) et massicoté. Cette dernière opération permit une numérisation de masse[89] utilisant un scanner recto-verso à chargeur[90].

A la sortie du scanner, nous disposions de répertoires contenant des images à haute définition compressées sans pertes[91], numérotées automatiquement. En se basant sur la description des volumes, nous pûmes automatiquement[92] produire :

- un rapport permettant de contrôler que le nombre de pages numérisées était cohérent avec la description des volumes,
- des archives, sur différents supports[93], où chaque image brute était renommée en fonction des méta-données du volume,

[89] Les dépliants, quant à eux, durent être numérisés « à la main ».

[90] Xerox Digipath.

[91] TIFF, compression CCITT Group 4, 600 points par pouce.

[92] Grâce au « Robot Transvision », logiciel développé à la MOM.

- des images pour la diffusion sur le Web (à une définition inférieure et compressées avec pertes[94]) ainsi que des vignettes.

La diffusion sur la Toile [Benel02b] est rendue possible à l'aide de deux types de serveurs HTTP. Le premier[95] permet de stocker les fac-similés et de les redimensionner en fonction des besoins de l'usager (taille de son écran). Tandis que le second[96] génère l'hypertexte permettant de feuilleter ces fac-similés. Notons que contrairement à ce que permettrait *Porphyre,* cet hypertexte n'autorise la navigation que dans une facette à la fois.

CEFAEL est hébergé au CINES, et profite donc de la puissance des machines du centre, de son réseau très haut-débit (nœud régional RENATER), et surtout de son équipe disponible 24h/24, 7j/7.

Pour conclure cette section, notons que la chaîne de production ainsi décrite permit d'atteindre, avec deux personnes affectées à la description et à la numérisation des ouvrages, une productivité de 40.000 pages par mois [Iacovella 2002].

3. Expérimentation dans Porphyre

Notre expérimentation eut lieu au cours de l'été 2001, au moment où seul un petit corpus de test avait été numérisé et était disponible dans l'intranet de l'EFA. La description du corpus fut exportée de la base de données vers *Porphyre.* Pour ce faire, nous dûmes définir un format d'échange pour les réseaux de description (à l'aide d'une DTD[97]), réaliser une petite « moulinette » pour générer le fichier correspondant à la base de données du corpus, ainsi qu'ajouter à *Porphyre* un module d'import pour ce type de fichiers.

---

[93] Notons tout de même que le nombre de CD-ROMs nécessaires à l'archivage du corpus est de l'ordre de 250 !

[94] JPEG, niveaux de gris, 150 points par pouce.

[95] Utilisant le système Transvision® développé par la MOM.

[96] Serveur « web » (Apache) agrémenté de scripts (développés en PHP) et d'une base de donnée (Sybase).

[97] Définition de type de document XML.

L'exploitation dans *Porphyre* de ce corpus de test fit l'objet d'une démonstration (cf. Figure 10.6) aux Journées Bibliothèques Numériques de mai 2002. Par ailleurs, en important un grand nombre de fois les mêmes volumes, nous pûmes tester la montée en charge des serveurs.

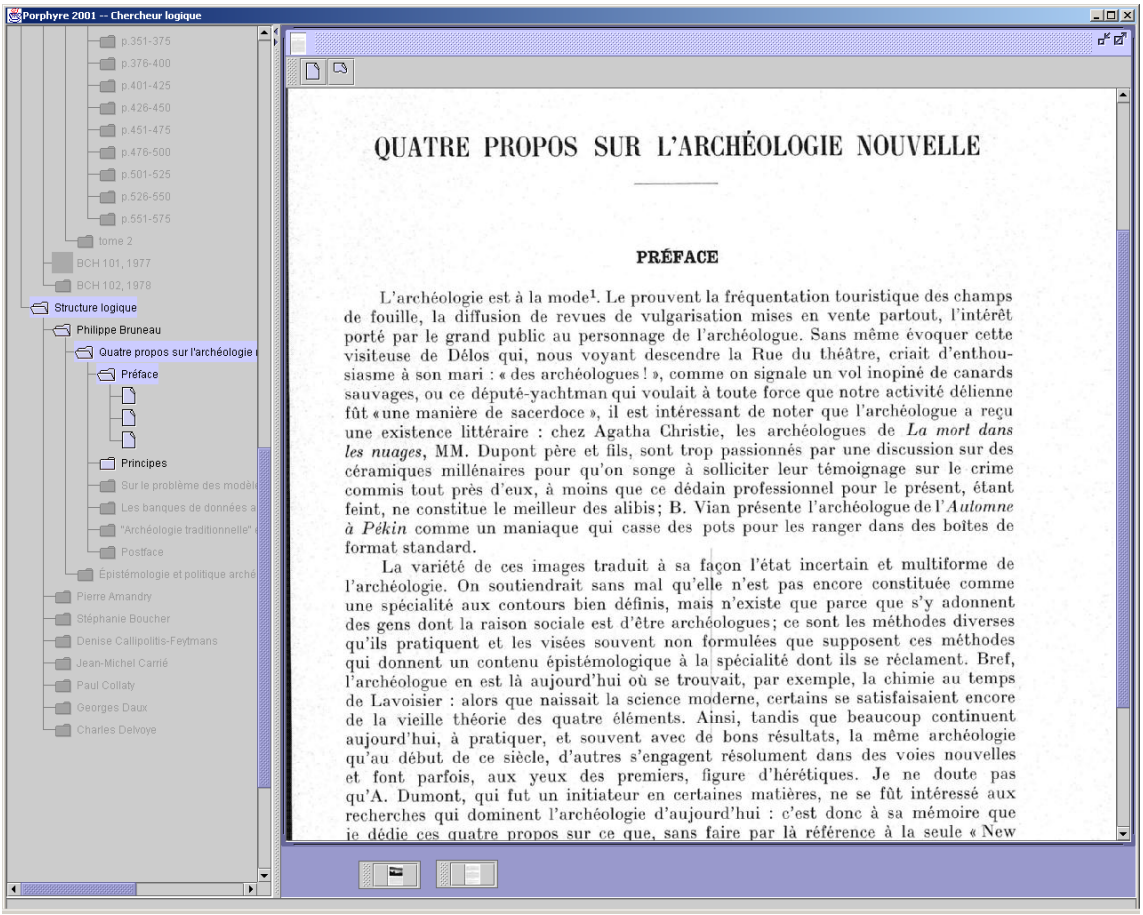

Porphyre 2001 -- Chercheur logique
p.351-375
p.376-400
p.401-425
p.426-450
p.451-475
p.476-500
p.501-525
p.526-550
p.551-575
tome 2
BCH 101, 1977
BCH 102, 1978
Structure logique
Philippe Bruneau
Quatre propos sur l'archéologie
Préface
Principes
Sur le problème des modèl
Les banques de données a
"Archéologie traditionnelle"
Postface
Épistémologie et politique arché
Pierre Amandry
Stéphanie Boucher
Denise Callipolitis-Feytmans
Jean-Michel Carrié
Paul Collaty
Georges Daux
Charles Delvoye
QUATRE PROPOS SUR L'ARCHÉOLOGIE NOUVELLE
PRÉFACE
L'archéologie est à la mode[1]. Le prouvent la fréquentation touristique des champs de fouille, la diffusion de revues de vulgarisation mises en vente partout, l'intérêt porté par le grand public au personnage de l'archéologue. Sans même évoquer cette visiteuse de Délos qui, nous voyant descendre la Rue du théâtre, criait d'enthousiasme à son mari : « des archéologues ! », comme on signale un vol inopiné de canards sauvages, ou ce député-yachtman qui voulait à toute force que notre activité délienne fût « une manière de sacerdoce », il est intéressant de noter que l'archéologue a reçu une existence littéraire : chez Agatha Christie, les archéologues de *La mort dans les nuages*, MM. Dupont père et fils, sont trop passionnés par une discussion sur des céramiques millénaires pour qu'on songe à solliciter leur témoignage sur le crime commis tout près d'eux, à moins que ce dédain professionnel pour le présent, étant feint, ne constitue le meilleur des alibis; B. Vian présente l'archéologue de l'*Automne à Pékin* comme un maniaque qui casse des pots pour les ranger dans des boîtes de format standard.
La variété de ces images traduit à sa façon l'état incertain et multiforme de l'archéologie. On soutiendrait sans mal qu'elle n'est pas encore constituée comme une spécialité aux contours bien définis, mais n'existe que parce que s'y adonnent des gens dont la raison sociale est d'être archéologues; ce sont les méthodes diverses qu'ils pratiquent et les visées souvent non formulées que supposent ces méthodes qui donnent un contenu épistémologique à la spécialité dont ils se réclament. Bref, l'archéologue en est là aujourd'hui où se trouvait, par exemple, la chimie au temps de Lavoisier : alors que naissait la science moderne, certains se satisfaisaient encore de la vieille théorie des quatre éléments. Ainsi, tandis que beaucoup continuent aujourd'hui, à pratiquer, et souvent avec de bons résultats, la même archéologie qu'au début de ce siècle, d'autres s'engagent résolument dans des voies nouvelles et font parfois, aux yeux des premiers, figure d'hérétiques. Je ne doute pas qu'A. Dumont, qui fut un initiateur en certaines matières, ne se fût intéressé aux recherches qui dominent l'archéologie d'aujourd'hui : c'est donc à sa mémoire que je dédie ces quatre propos sur ce que, sans faire par là référence à la seule « New

Figure 10.6 : Lecture avec *Porphyre 2001* d'un extrait des *Collections de l'École française d'Athènes en ligne*

4. Retour d'expérience

Le prototype d'alors était basé sur un serveur de contenu et un serveur de structure. Le premier était constitué de « servlets » appliquant des algorithmes « maison » à des images JPEG. Le second était conçu de telle sorte que les réseaux de description puissent dépendre les uns des autres suivant un ordre partiel. Par exemple, tout descripteur du réseau d'une bibliothèque pouvait être généralisé par un descripteur du réseau

d'un chercheur, à condition que ce chercheur soit « abonné » à la bibliothèque. Chaque serveur devenait alors le client de plusieurs autres.

Le premier problème rencontré concernait le serveur de contenu. D'une part, il était regrettable de ne pouvoir gérer que des versions dégradées (JPEG) des fac-similés. Ensuite, la performance de l'architecture à base de servlets et d'algorithmes « maisons » s'est avérée insuffisante. La nouvelle version à base de scripts PHP, intégrant des composants externes optimisés, a permis un gain de performance considérable [Tribollet03].

Le second problème concernait le serveur de structure. Le mode de distribution des données ne permettait de tirer aucun profit de la mise en parallèle des calculs sur les différents serveurs. Pour remédier à cela, nous avons défini les notions d'objets documentaires et de facettes. Aujourd'hui deux réseaux de description ne dépendent l'un de l'autre que par l'intermédiaire des objets documentaires. Au niveau de l'architecture, le client interroge directement les serveurs. L'intégration des données est rendue possible par le fait que les serveurs se réfèrent aux mêmes serveurs de correspondance. Avec la nouvelle architecture, si l'on gère *n* facettes sur *n* serveurs différents la charge des serveurs sera *n* fois moindre que sur un serveur unique. Une autre optimisation est également envisagée. Elle consisterait à tirer partie du fait que la facette du maquettiste est arborescente. Le filtre étant beaucoup moins complexe à calculer avec de telles structures, il serait judicieux de développer un serveur spécialisé implémentant le même protocole mais de manière optimisée.

Chapitre 11. La nécropole occidentale de Mégara Hyblaea

Dans les études de cas précédentes, les réseaux de description ont été utilisés pour exprimer le point de vue des auteurs, des éditeurs ou des bibliothécaires. Pourquoi ne pas avoir traité le point de vue des lecteurs, celui que nous avions présenté comme le plus intéressant ? En fait, une telle expérimentation pouvait difficilement commencer avant que nos modèles théoriques et notre prototype n'aient atteint une certaine maturité. Il fallut attendre cette année pour qu'une équipe pilote d'archéologues (cf. Chapitre 1) soit intégrée au projet et formée à l'usage de *Porphyre*. Les premières expérimentations pourront commencer à l'automne 2003. En attendant ces retours d'expérience, nous avons souhaité éprouver notre prototype en nous mettant « à la place » d'un archéologue. Nous sommes conscients de la portée très relative d'une telle expérience. Toutefois, il nous semble que les défauts du modèle qui pourraient apparaître dans notre usage de *Porphyre* devraient *a fortiori* causer des problèmes aux archéologues et avoir ainsi valeur de réfutation.

Cette étude de cas portera sur les recherches d'Andrea Iacovella concernant la nécropole occidentale de Mégara Hyblaea (Sicile). Dans une première partie, nous essaierons de nous mettre dans la situation du chercheur en présentant ses objectifs et méthodes. Dans une deuxième partie, nous verrons les problèmes rencontrés avec la précédente version de *Porphyre* et surtout la difficulté de faire une description avancée sans être guidé. Dans une troisième partie, nous esquisserons par conséquent ce qui pourrait devenir à terme un « guide des bons usages ».

1. « Fouiller » un rapport de fouille

A la croisée de l'archéologie, de l'historiographie[98] et des sciences cognitives, les travaux d'Andrea Iacovella visent à analyser le discours des archéologues [OrsiEt-Cavallari1892] afin d'en extraire de nouvelles conclusions archéologiques. En quelque sorte, il s'agit de refaire, virtuellement, une fouille effectuée au siècle dernier.

---

[98] Historiographie : Etude de l'écriture de l'Histoire.

Après avoir développé dans sa thèse une approche quantitative (à l'aide de statistiques descriptives), Andrea Iacovella souhaitait passer à une approche plus qualitative. Dans cette perspective, François Dubois (projet de fin d'études INSA) effectua une « modélisation cognitive » d'une vingtaine de descriptions de sépultures [Dubois99]. Ses modèles prenaient la forme de diagrammes d'instance UML étiquetés avec les termes de Paolo Orsi (traduits de l'Italien en Français). Durant l'été 2002, nous intégrâmes dans *Porphyre* le rapport de fouille (sous forme de fac-similés de pages) et « traduisîmes » les modèles UML en un réseau de description.

Nous considérâmes trois facettes pour décrire le corpus : une première pour sa structure typographique en colonne, une seconde pour sa structure par sépulture et une troisième pour la typologie des vestiges trouvés dans ces sépultures. Volontairement, nous adoptâmes un modèle simple dans lequel la composition de descripteurs ne se faisait qu'au niveau du fragment (intersection du découpage par colonne et par sépulture) et où les adjectifs n'étaient pas pris en compte. Nous plaçant dans la perspective d'une modélisation dynamique, il semblait en effet naturel de commencer par des modèles « naïfs » et de les affiner par la suite.

2. Retour d'expérience

a. Premiers résultats

Le filtre de graphe appliqué à notre réseau de description permit d'observer un certain nombre de propriétés en résonance avec les préoccupations d'Andrea Iacovella. Dans la Figure 11.1, par exemple, il apparaît que, pour le corpus décrit, les sépultures d'enfants contiennent toutes des lécythes (mais jamais en forme de cœur) et qu'aucune ne contient de masque féminin ou d'aiguille. L'archéologue aurait alors pu se demander si le matériel de la sépulture est déterminé par l'age du défunt. A l'inverse, l'historiographe, aurait pu se demander si, dans le cas de crémations (donc en l'absence de squelette), ce n'est pas le matériel qui permit au fouilleur de déterminer l'age du défunt.

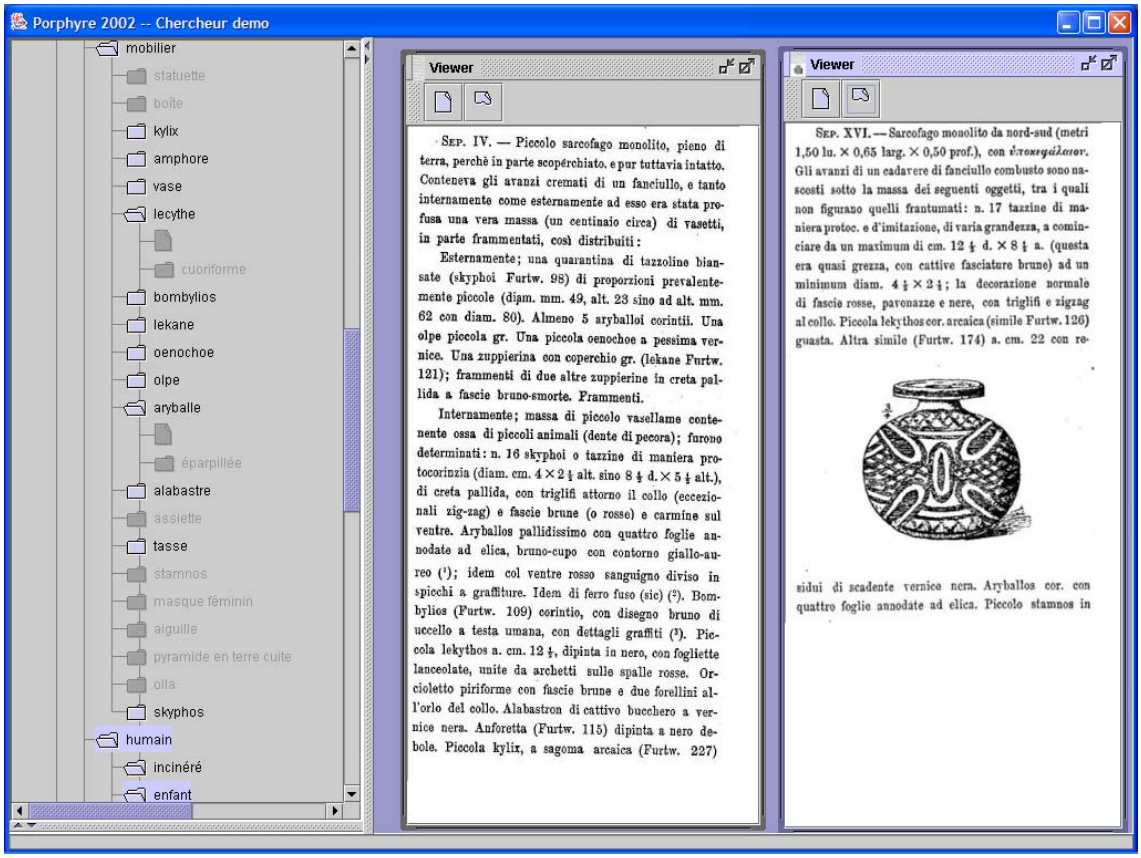


Figure 11.1 : Lecture active avec *Porphyre 2002* d'un extrait de la publication des fouilles de la nécropole de Mégara Hyblaea

b. Et au-delà ?

Pour approfondir cette étude, l'archéologue pourrait s'intéresser à la répartition spatiale des tombes d'enfants par rapport aux tombes d'adultes (cf. [Iacovella97]). Il faut alors ajouter au corpus une carte de la nécropole, créer un fragment pour chaque emplacement de sépulture, et associer ce fragment au descripteur correspondant à la sépulture. Toutes les données permettant d'afficher la carte de répartition des tombes d'enfants sont alors disponibles. Mais, nous nous retrouvons ici face à une limite de la version 2002 de *Porphyre* qui ne permet pas d'afficher sur un document source les emplacements de plusieurs fragments. Dans la version 2003 en cours de développement, la refonte du serveur de contenu ainsi que l'introduction du serveur de correspondance (cf. Chapitre 6) permet, de manière élégante, de dépasser cette limitation.

Continuons à nous mettre à la place de l'archéologue. Celui-ci pourrait par exemple se demander si la petite taille du matériel est, elle aussi, corrélée avec le jeune âge du défunt. Mais comment modéliser l'adjectif « petit » ? Nous sommes bien au-delà de la modélisation naïve que nous préconisions au début de ce chapitre. Il paraît donc nécessaire d'offrir aux experts qui le souhaitent une formation avancée sur la description de documents. La section suivante est une esquisse de ce que pourrait être une telle formation.

3. Esquisse d'un « guide des bons usages »

a. Modéliser une taxinomie

La relation de spécialisation entre le descripteur *A* et *B* peut être utilisée pour exprimer que « *A* subsume *B* ». C'est-à-dire que *A* et *B* modélisent des classes pour lesquelles toute instance de *B* est aussi une instance de *A*, et toute sous-classe de *B* est une sous-classe de *A*. Par exemple (cf. Figure 11.2), la classe « Coiffure » subsume les classes « Polos » et « Couronne ».

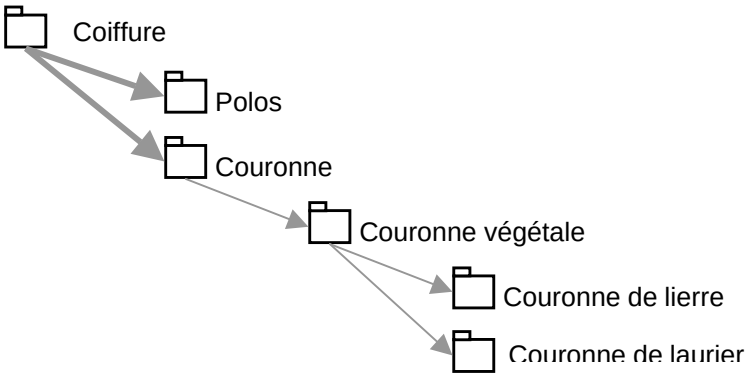


Figure 11.2 : Modélisation d'une taxinomie à l'aide des réseaux de description

On peut si on le souhaite définir une classe « Bombylios corinthien » par composition de « Bombylios » et de « Vestige corinthien », cependant, cela complique la gestion du réseau (cf. « Petit bombylios» de la Figure 11.3). Il est donc conseillé d'effectuer cette composition au niveau des instances et non des classes.

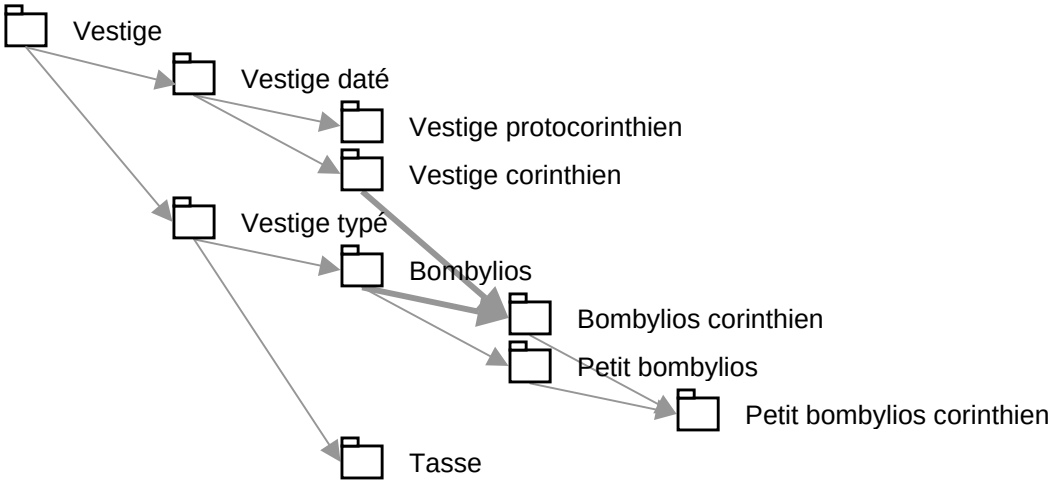


Figure 11.3 : Taxinomie avec composition : à utiliser avec discernement

b. Modéliser des attributs et des valeurs d'attribut

Une autre manière de modéliser l'exemple précédent est de considérer les relations de *A* vers *B* et de *B* vers *C* comme voulant dire « *B* est un attribut de *A* et peut prendre *C* pour valeur». Ainsi (cf. Figure 11.4), « Style » est un attribut de « Vestige » et « Protocorinthien » est l'une de ses valeurs possibles.

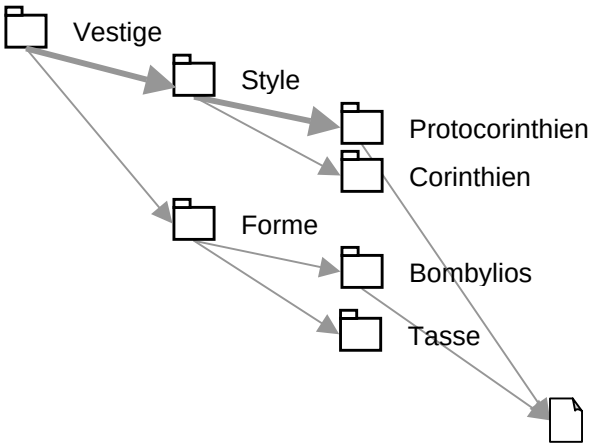


Figure 11.4 : Modélisation d'attributs à l'aide des réseaux de description

c. Modéliser une méréonymie

La relation de spécialisation entre le descripteur *A* et *B* peut être utilisée pour exprimer que « l'objet *B* est une partie de l'objet *A*».

On utilisera la méréonymie pour exprimer par exemple les relations entre les sections d'un document, ou bien les relations d'inclusion entre zones spatiales ou même entre intervalles temporels.

Notons que la polyhiérarchie est ici particulièrement utile et permettra par exemple d'exprimer (cf. Figure 11.5) que « l'Age du Fer » empiète sur les « Temps protohistoriques » et les « Temps historiques » (intersection non vide). Et que la « Période romaine » appartient à la fois à « l'Age du Fer » et aux « Temps historiques ».

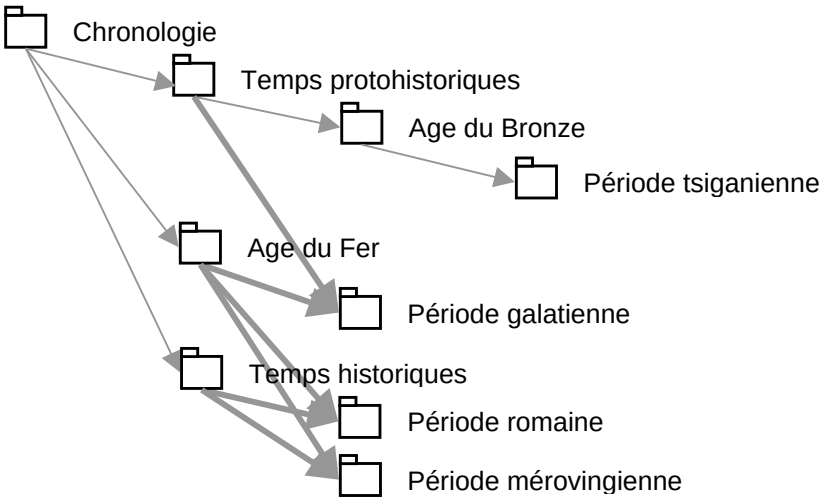


Figure 11.5 : Méréonymie utilisée pour décrire une chronologie (celle de G. de Mortillet [Jockey99])

d. Modéliser des instanciations

La relation de *A* à *B* peut également être utilisée pour modéliser que « l'objet *B* est une instance de la classe *A* ». Par exemple (cf. Figure 11.6), « un petit bombylios corinthien en mauvais état » est une instance de la classe « Bombylios ». Notons que la polyhiérarchie nous permet de préciser que cet objet est aussi une instance de « Vestige corinthien ».

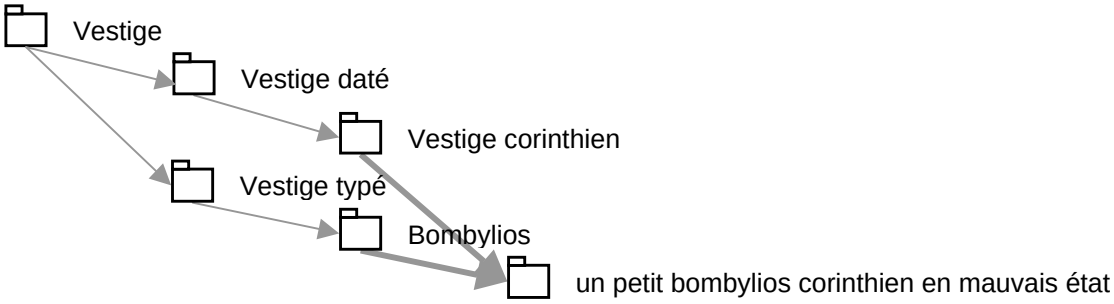


Figure 11.6 : Instanciation multiple

e. Modéliser des liens d'association non orientés

Pour modéliser un lien d'association (au sens UML) entre deux objets *A* et *B*, nous utiliserons un descripteur *C*, spécialisation de *A* et de *B.* Ainsi (cf. Figure 11.7), les objets « un squelette » et « un vase » sont-ils reliés par le lien « sont au même niveau ».

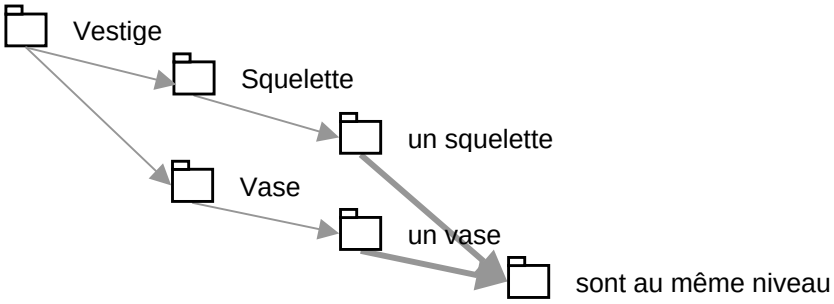


Figure 11.7 : Lien d'association « sont au même niveau »

f. Modéliser des associations non-orientées

On peut également souhaiter typer un lien d'association (par exemple pour rechercher les liens de même type). Nous devons alors généraliser le lien par un descripteur qui jouera le rôle d'une association. Ainsi, le réseau de la Figure 11.8 modélise-t-il le fait que « deux squelettes sont tête-bêche », association qui pourra être réutilisée dans d'autres descriptions.

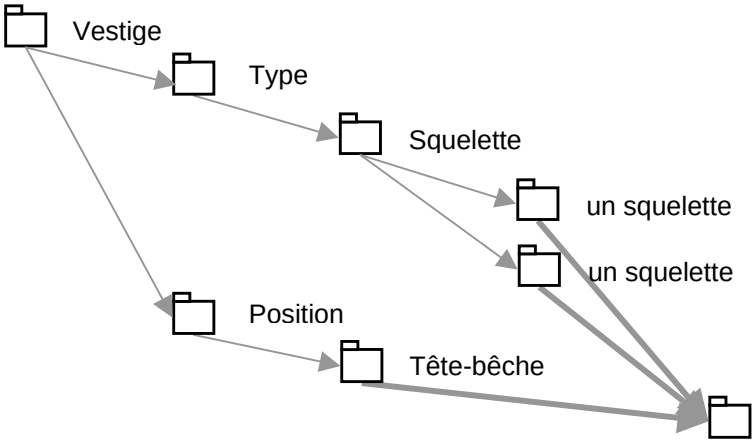


Figure 11.8 : Association non-orientée « Tête-bêche »

g. Modéliser des associations orientées

Pour modéliser une lien d'une association orientée, il faut créer un descripteur pour chaque rôle, faire une spécialisation commune à chaque rôle et à l'objet jouant ce rôle, puis faire une spécialisation commune à ces derniers descripteurs. Ainsi, la figure 11.9 exprime-t-elle qu'un sac contient un vase, tout en précisant que les rôles « contenu » et « contenant » sont réutilisables (on pourrait par exemple rechercher tous les objets contenus dans les vases).

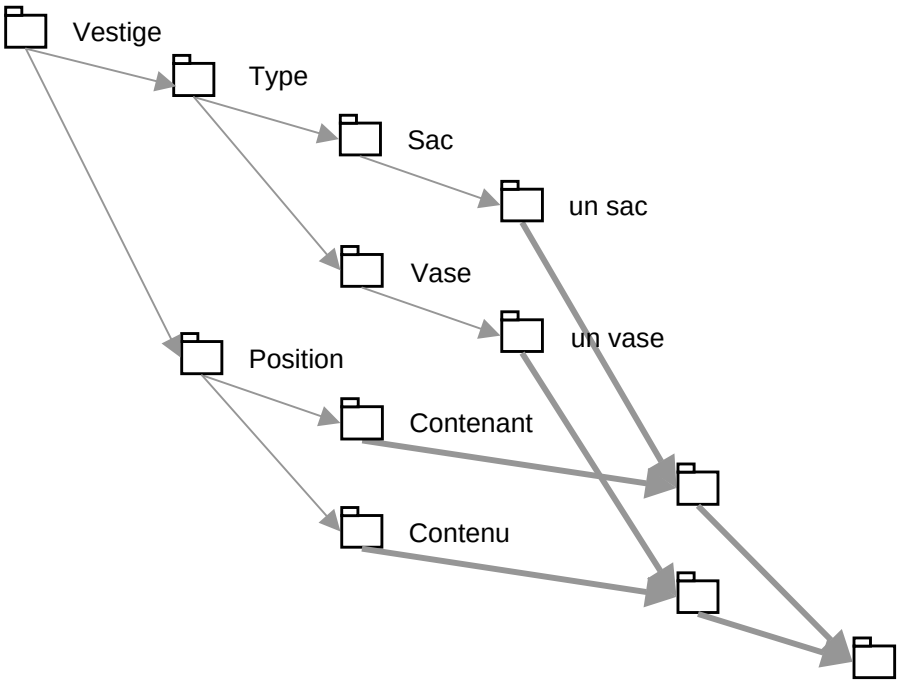


Figure 11.9 : Association orientée « contient »

h. Utiliser des documents comme « points de repère »

Si les réseaux de description aident à la compréhension des objets documentaires, en retour les objets documentaires permettent d'interpréter les réseaux de description. En effet, qu'y a-t-il de mieux pour savoir ce qu'est qu'un bombylios que de voir les documents textuels et graphiques qui ont été décrits par ce descripteur ?

On peut étendre cette pratique en ajoutant des documents prévus spécifiquement pour illustrer les descripteurs : définitions textuelles, figures représentant des exemples typiques, cartes géographiques, frises chronologiques, introduction, table des matières, table des figures, etc. Placés dans le réseau de sorte qu'ils dépendent directement du (ou des) descripteur(s) à illustrer, ils seront affichés par le filtre comme autant de « points de repères » qui aideront le lecteur à s'orienter.

Chapitre 12. Perspectives

En fournissant des outils pour la lecture, l'écriture et la publication de corpus, nous avons souhaité que des pairs clairement identifiés puissent donner un avis critique sur les documents[99]. Le système *Porphyre* comprend aujourd'hui deux couches (au sens informatique) : la première permettant de gérer les corpus documentaires, la seconde les points de vue. Il est donc devenu un espace de débat dans lequel chacun peut superposer au corpus sa propre lecture. Dès lors, il serait souhaitable d'envisager la création de deux couches supplémentaires : une pour l'intersubjectivité et l'autre pour la diachronie.

Après avoir précisé ce que nous entendons par « espace intersubjectif » et « espace diachronique », nous présenterons deux champs d'application, celui du temps archéologique et celui des documents d'architecture en archéologie.

1. Espace intersubjectif[100]

a. principe

L'espace intersubjectif permettra de détecter des incohérences par propagation de contraintes, contraintes posées par l'expert à l'intérieur d'un même point de vue et entre points de vue différents.

Précisons que vérifier la cohérence des modèles produits par l'ensemble de la discipline aurait d'une part un coût algorithmique exorbitant et serait d'autre part d'une utilité très limitée. Aussi, nous préférons offrir à chaque expert des espaces dans lesquels ils pourront importer les parties de points de vue qu'ils souhaitent confronter.

Parce que la couche intersubjective s'appuiera sur les deux précédentes, il sera possible à tout moment de connaître pour une partie de point de vue sa situation d'origine (Qui l'a créé, modifié ? Quand ? Pour quelle communauté ?...) ainsi que son

---

[99] Condition d'ailleurs présentée par R. H. Tibbo [Tibbo93] comme nécessaire à la recherche d'information en Sciences Humaines.

[100] Nous défendons actuellement un dossier d'identification pour une équipe-projet STIC sur ce thème.

contexte d'origine (point de vue dans sa globalité et corpus documentaire sur lequel il s'appuie).

b. Assistance à la pose de contraintes inter- points de vue

La détection d'incohérences entre points de vue n'est possible que si on ajoute des relations (égalité, subsomption, séquence, etc.) entre des descripteurs appartenant à des points de vue différents.

Par exemple, pour comparer deux structurations de corpus, un expert pourra affirmer que le site archéologique « Shisma Eloundas » est équivalent au site appelé par un autre auteur « Schisma » partie de « la région d'Elounda » (car dans « Eloundas » le « s » est la marque d'un génitif grec). Autre exemple, l'expert pourra faire une étude comparative des datations de différents sites en explicitant le fait que lorsqu'un auteur parle de l'époque « MR III », il s'agit en fait d'une subdivision de ce qu'un deuxième auteur appelle « Minoen récent ».

S'il est clair que de telles correspondances ne peuvent être faites que par les experts eux-mêmes, il faudrait toutefois les assister si l'on veut comparer des modèles de taille importante. Nous proposons de leur fournir une sorte de moteur de « macros » leur permettant de rechercher automatiquement certains « patrons » qu'ils définiront puis d'ajouter une contrainte donnée si la correspondance est validée par l'expert.

Pour reprendre nos exemples, l'expert choisira de rechercher tous les couples de descripteurs (*A, B*), descendants d'un descripteur ayant pour étiquette « Région », et pour lesquels *A* a une étiquette de la forme « *X Y*s » et *B* a pour étiquette « *X* » et pour parent un descripteur ayant pour étiquette « *Y* ». Dans ces cas-là, le système proposera une contrainte d'équivalence entre *A* et *B* (nous pourrons ici développer par exemple les travaux de [MatthesEtAl01]).

c. Consensus et importation

Dans le cas où les experts arriveront à régler le conflit d'interprétation qui les opposait, il sera possible de publier le résultat. En pratique, la fusion de leurs points de

vue (conformément à des règles à définir), pourra être importée dans un point de vue propre à la communauté. Le système gardera la mémoire de l'auteur originel de chaque partie de sorte que la copie soit une citation et non un plagiat.

2. Espace diachronique

A la différence des autres espaces dédiés à la production, l'espace diachronique sera consacré à leur observation. Il s'agira de visualiser l'activité scientifique en montrant les variations qui ont affecté le domaine. L'analyse portera sur l'évolution du vocabulaire de la communauté, celle de leur emploi pour décrire le corpus documentaire, et sur l'influence réciproque dans le temps des modèles de la communauté et de ceux de ses membres.

Si *Porphyre* permet dès aujourd'hui de capter les informations historiques concernant les traces d'interprétation stockées, il s'agit d'aller plus loin et de trouver des moyens de présenter visuellement et synthétiquement l'évolution des traces en contexte. Notons que ceci représente un double défi dans le domaine des bases de données visuelles et celui des bases de données temporelles.

3. Application au temps archéologique

a. Présentation

Le travail de thèse de Tiphaine Accary, commencé à l'automne 2002, s'inscrit dans la perspective intersubjective que nous avons présentée. Les règles permettant de détecter des contradictions entre points de vue étant dépendantes d'une discipline et d'une méthode, un champ d'application a été choisi : le temps archéologique[101].

b. « Drôle de temps »

Pour comprendre l'intérêt que représente une telle orientation de recherche, il est nécessaire d'en dire un peu plus sur le temps archéologique. Celui-ci est souvent présen-

[101] Ce travail fait suite à notre collaboration avec Nicolas Gorse [Gorse03].

té comme l'une des trois dimensions principales de l'archéologie avec l'espace et les types de vestiges. Cependant, l'expression « dimension » n'est aucunement à prendre au sens mathématique. Autrement dit, le temps est dépendant de l'espace et des vestiges.

Comme le rappellent Patrick Desfarges et Bruno Helly [DesfargesEtHelly91] : « contrairement à une opinion bien ancrée dans l'esprit des archéologues, le temps, la chronologie n'est pas une donnée mais un résultat ». Le temps archéologique est donc construit à partir d'indices spatiaux, stylistiques ou naturels[102] en utilisant des heuristiques plus ou moins fiables. Cette construction a pour but, selon Alain Gallay [Gallay86], de rechercher un ordre pour saisir l'événement et le changement. A la suite de J. P. Demoule [Demoule72], il faudra tenir compte du fait qu'en archéologie il n'existe pas une chronologie unique mais des chronologies valables dans des régions géographiques données, avec parfois des correspondances possibles entre elles.

c. Travaux actuels et à venir

La doctorante, en partant de la caractérisation du temps archéologique qu'en font les théoriciens de la discipline, tente actuellement d'établir un état des principales « écoles méthodologiques » en ce qui concerne la formalisation du temps archéologique. Parallèlement, elle étudie en quoi les modèles temporels de l'intelligence artificielle (par exemple celui de James F. Allen [Allen83]) correspondent ou non aux spécificités de l'archéologie. Ses premiers résultats [Accary02] ont fait l'objet d'une publication [AccaryEtAl03].

Son but est de mettre en œuvre un assistant permettant entre autres de détecter les cas où une fouille vient contredire la chronologie de référence. En filigrane, se pose la question de la réfutabilité des théories en archéologie, enjeu majeur pour la discipline.

A des fins de validation, cet assistant sera intégré au système *Porphyre* pour être utilisé dans le cadre de bibliothèques numériques réelles.

---

[102] Au sens de physique ou chimique.

4. Application au document d'architecture (en archéologie)

a. Présentation

Le sujet de thèse de Florent Ruard-Dumaine (début de la thèse prévu pour la rentrée universitaire 2003) porte sur les rapports entre sémiotique graphique, sémantique lexicale et interprétation. Cette thématique, autour de la question de la construction du sens, sera étudiée dans le cadre du document d'architecture en archéologie. Notons que cette thèse sera aussi une thèse alliant informatique et archéologie mais cette fois-ci avec une dominante archéologique. Le doctorant sera intégré à une équipe interdisciplinaire[103].

L'architecture antique étant relativement normée, le doctorant pourra étudier pour un type d'ensembles architecturaux (par exemple les thermes), les règles de composition en usage, afin de dresser une sorte de « grammaire » architecturale (cf. Figure 12.1a). En particulier, il fera le point sur les outils lexicologiques qui constituent un savoir de référence partagé par une même communauté de spécialistes (dictionnaires et thésaurus).

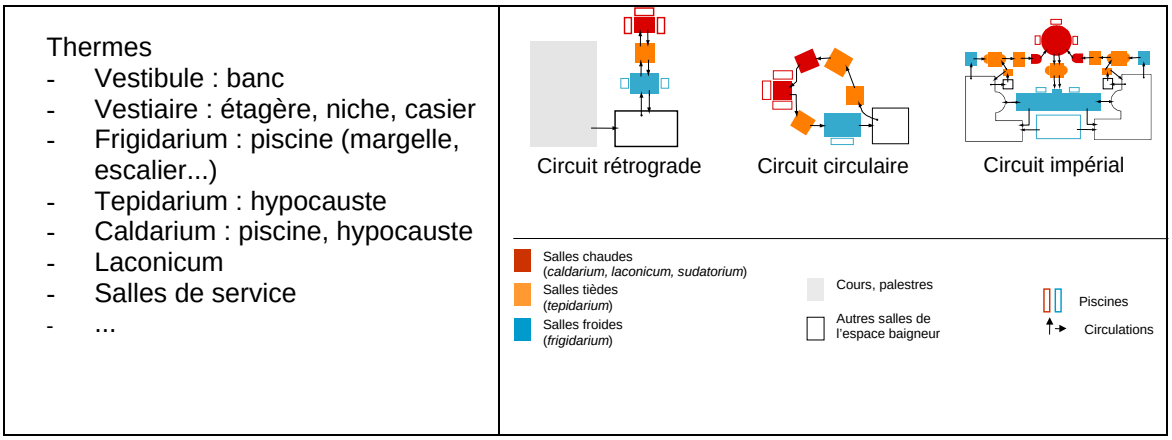


Figure 12.1 : Exemple de savoir de référence : structure du lexique[104] et structure du document[105]

[103] Action Concertée Incitative rassemblant des archéologues, des architectes et des informaticiens de l'EFA, du LIRIS et de la Maison de l'Orient et de la Méditerranée.

[104] D'après le *dictionnaire méthodique de l'architecture grecque et romaine* [GinouvesEtAl98].

[105] Th. Fournet d'après D. Krenker et I. Nielsen.

Parallèlement, toujours pour le même type d'ensembles architecturaux, le doctorant établira un modèle sémiotique associant à la structure du document (suivant son type) l'organisation spatiale des bâtiments (cf. Figure 12.1b). En effet, qu'il s'agisse de documents archéologiques primaires (photographies, relevés topographiques, croquis…) ou de documents d'architecture finaux (relevés pierre à pierre, plans masses, élévations, axonométries…), leur sémiotique spatiale est omniprésente.

Enfin, le doctorant étudiera la contribution du document d'architecture à la discipline archéologique. Il dressera un état des pratiques d'exploitation et d'interprétation de ces documents par les experts, en particulier leurs modes de lecture et d'annotation (cf. Figure 12.2). Le résultat ultime consistera à proposer des méthodes semi-automatiques, à intégrer dans *Porphyre*, permettant de détecter les incohérences entre le savoir de référence et celui de l'expert.

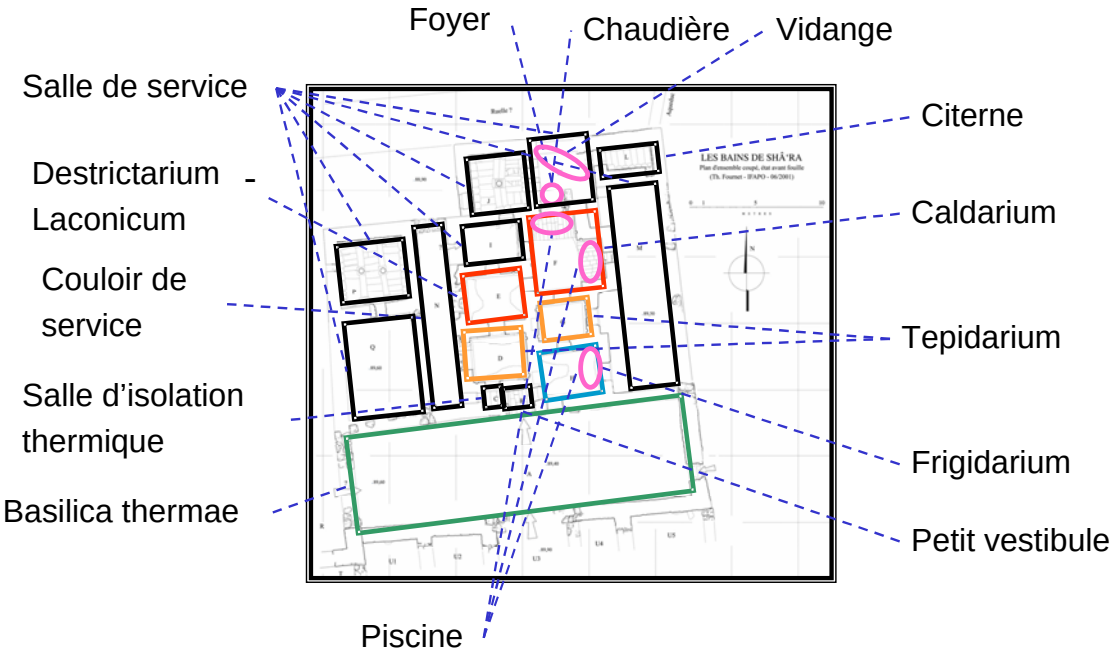


Figure 12.2 : Exemple d'annotation d'un document d'architecture[106] par un expert

Du point de vue informatique, l'enjeu principal porte sur la gestion conjointe (modification, comparaison, navigation, interrogation) de deux types de modèles for-

[106] Plan de Th. Fournet.

mels : réseaux sémantiques et profils 2D, étudiés habituellement par deux communautés distinctes. L'enjeu sera d'autant plus fort que le modèle de référence sera dynamique et hypothétique.

Du point de vue des Sciences Humaines, ce travail de thèse permettra de préciser les liens existant entre sémiotique graphique et sémantique lexicale. Il permettra également d'explorer la question du rapport au document dans les disciplines historiques. Enfin, il posera la question du statut épistémologique de l'espace archéologique, à savoir : « Est-ce une donnée ou une construction ? ».

Conclusion

En introduction, nous proposions d'explorer la question de l'instrumentation informatique de la consultation de corpus en Sciences Humaines. Que pouvons nous en dire à l'issue des douze chapitres de cette thèse ?

(1) Face aux besoins de l'Ecole française d'Athènes en termes d'accès à ses corpus numérisés, un certain nombre de réponses existent dans les domaines de la Recherche d'Information, de la Modélisation des Connaissances et de l'Hypertexte. Cependant pour chacune de ces approches, subsistent des critiques fondamentales portant non sur les techniques elles-mêmes mais sur le cadre philosophique dans lequel elles sont utilisées.

(2) Sachant que notre projet s'adresse à des scientifiques, dans quel cadre philosophique devons-nous nous placer ? Autrement dit, quelle valeur de connaissance donner aux modèles informatiques d'un corpus scientifique ? Il appert que, pour être en adéquation avec l'épistémologie contemporaine, nous devons considérer ces modèles comme la formalisation d'un système d'hypothèses, matière à débat et à évolution.

(3) On est en droit de se demander, ensuite, si le fait que notre cadre d'application soit celui des Sciences Humaines ne nous oblige pas à prendre en compte un certain nombre de particularités épistémologiques. En fait, en rattachant les Sciences Humaines à la tradition herméneutique[107], il apparaît indispensable de replacer les textes dans leur contexte (linguistique) et leur situation (extralinguistique). De ceci, dérivent deux notions clefs à instrumenter : celle de l'intertextualité et celle de l'intersubjectivité. La première indique que deux textes lus en parallèle sont d'avantage porteurs de sens que lus séparément, comme s'ils sélectionnaient du sens l'un dans l'autre. La seconde repose sur l'idée que le sens peut naître des conflits d'interprétation entre lecteurs.

(4) Si ces deux notions ressortent clairement des enquêtes portant sur les usages dans les bibliothèques traditionnelles, peu de bibliothèques numériques en tiennent

[107] Herméneutique (adj.) : Relatif à l'interprétation des textes.

compte aujourd'hui. Le domaine des hypermédias, quant à lui, a placé l'intertextualité au cœur de son approche, mais on est loin de l'intersubjectivité. En effet, il est souvent fort difficile d'exprimer des points de vue différents sur un même objet documentaire.

(5) Le modèle hypermédia que nous introduisons, appelé *Porphyre*, consiste à capter le geste du lecteur. Chaque manipulation du corpus (annotation, mise en corrélation, commentaire…) est alors considérée comme un *trace* d'interprétation à archiver. Par cette stratégie de lecture/écriture, l'expert arpente l'espace documentaire afin d'en faciliter l'accès ultérieur (pour lui-même ou, s'il le souhaite, pour sa communauté).

(6) Les *contenus documentaires* constituent notre premier type de trace. Le modèle permet de gérer des fragments, des sources, des notes dynamiques ainsi que des contextes de lecture (plusieurs objets documentaires destinés à être lus conjointement). En tant que référence commune d'une communauté, ces contenus documentaires constituent en quelque sorte l'assise du modèle.

(7) Notre deuxième type de trace est constitué par les *réseaux de description*. Ils permettent à chaque expert de superposer au corpus sa propre structure, son propre point de vue. Un mécanisme de filtre (obtenu par induction totalisante) permet d'aider l'expert à arpenter l'espace documentaire conjointement à travers plusieurs points de vue.

(8) Les *parcours de lecture* forment notre troisième type de trace. Il s'agit de structures, orthogonales aux réseaux de description, permettant de définir (lors de la constitution du corpus ou lors de son exploration) des séquences à travers les objets documentaires. Chaque objet documentaire peut ainsi devenir l'embranchement de parcours appartenant à différents points de vue. De plus, à chaque étape de lecture, l'objet documentaire est associé à un contexte de lecture particulier.

(9) Le modèle étant défini, reste sa mise à l'épreuve. La première étude de cas, remontant à l'an 2000, concernait un extrait de la *Chronique des fouilles*. L'expérience visait à enrichir les documents, structurés par « livraison » (année de fouille), à l'aide d'une structure géographique générale (région, site) de manière à obtenir en quelque sorte une « topique » des fouilles. Les difficultés rencontrées nous incitâmes par la suite

à gérer les structures originelles des corpus comme des points de vue parmi d'autres. Par ailleurs, c'est cette étude de cas qui nous montra l'importance des séquences et nous mit ainsi sur la voie des parcours de lecture. Enfin, nous constatâmes la nécessité d'outils d'assistance à la fusion de structures.

(10) Notre deuxième étude de cas, remontant à 2001, portait sur la gestion avec *Porphyre* d'un extrait des Collections de l'Ecole française d'Athènes en ligne (CEFAEL). Nous prîmes en compte les structures suivantes : celles du maquettiste, de la bibliothèque, de la photothèque/planothèque, de l'équipe de fouille. Les problèmes rencontrés portèrent principalement sur le passage à l'échelle. Ceux-ci nous obligèrent à modifier des choix technologiques (servlets) ainsi que des choix d'architecture (communications entre serveurs).

(11) Notre troisième étude de cas, menée en 2002, portait sur une lecture historiographique de la publication d'une fouille de nécropole. Nous considérâmes les structures suivantes : la structure bibliographique du rapport (en colonne), la structure par sépulture et la typologie des vestiges (tombes, mobilier, restes humains). Cette expérimentation nous encouragea d'une part à revoir notre gestion des contextes de lecture (pour afficher par exemple une carte de répartition du matériel archéologique) et, d'autre part, à rédiger un « guide des bons usages » à l'intention des experts souhaitant créer des modèles complexes.

(12) Le modèle actuel de *Porphyre* laisse envisager un certain nombre d'évolutions. Une première consisterait en la création d'un espace intersubjectif permettant d'expliciter les relations entre points de vue et de faire ressortir ainsi les zones d'achoppement. Une seconde correspondrait à la constitution d'un espace diachronique offrant la possibilité de visualiser la dynamique des points de vue. Ces deux aspects pourront plus particulièrement être étudiés dans le cas de la manipulation par l'archéologue du document d'architecture et du temps archéologique.

Pour conclure, nous pourrions dire que si cette thèse ne prétend pas avoir épuisé la problématique initiale, nous pensons que sa clôture entame une nouvelle dynamique dans l'histoire du projet. Pour reprendre les définitions de Basarab Nicolescu [Nicoles-

cu96], nous serions en train de passer d'une dynamique *interdisciplinaire* à une dynamique *transdisciplinaire*.

En effet, notre point de départ consistait à transférer des méthodes d'une discipline à une autre dans une logique à la fois d'application et de questionnement des disciplines (de l'informatique à l'archéologie et en retour de l'épistémologie et de la linguistique à l'informatique).

Aujourd'hui, avec la constitution d'une équipe regroupant des chercheurs en archéologie, en historiographie, en architecture et en informatique, le but poursuivi est que chacun, à partir de sa discipline, s'interroge sur la question de la création du sens, thème qui traverse et dépasse toutes les disciplines.

Bibliographie*

[AccaryEtAl03] Accary T., Bénel A., Calabretto S., Modélisation de connaissances temporelles en Archéologie [en ligne], In : *Actes des Journées francophones d'Extraction et de Gestion des Connaissances [EGC'2003], Revue des Sciences et Technologies de l'Information [RSTI], 2003, Vol. 17, Numéro spécial*. Paris : Lavoisier – Hermès Sciences, 2003. p.503-508. Disponible sur Internet : <http://lisi.insa-lyon.fr/~taccary/publications/Time_EGC03.pdf> (consulté le 26/09/03)

[Accary02] Accary T., *Instrumenter le travail des archéologues : la modélisation du temps* [en ligne], DEA en Informatique, Université Lyon 1, 2002. Disponible sur Internet : <http://lisi.insa-lyon.fr/~taccary/publications/dea_memoire.zip> (consulté le 26/09/03)

[Ackerman94] Ackerman M.S., Providing Social Interaction in the Digital Library [en ligne], In: Digital Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, College Station (Texas), June 19-21 1994. Disponible sur Internet : <http://www.csdl.tamu.edu/csdl/DL94/position/ackerman.html> (consulté le 26/09/03)

[AitKaciEtAl89] Aït-Kaci H., Boyer R.S., Lincoln P., Nasr R., Efficient Implementation of Lattice Operations [en ligne], In: *ACM Transactions on Programming Languages and Systems, Vol. 11, No 1 (Jan. 1989).* p.115-146. Disponible sur Internet : <http://doi.acm.org/10.1145/59287.59293> (consulté le 26/09/03)

[Allen83] Allen J., Maintaining knowledge about temporal intervals [en ligne], In: *Communications of the ACM 26(11),* 1983. p.832-843. Disponible sur Internet : <http://doi.acm.org/10.1145/182.358434> (consulté le 26/09/03)

---

* Il nous a semblé plus pertinent d'un point de vue de l'histoire des sciences d'indiquer dans la clef de citation l'année de première édition, plutôt que celle de l'édition consultée. Dans la bibliographie, l'édition consultée est mentionnée dans le corps de la référence, tandis que l'édition originale est décrite en notes. Exemple :

[Dreyfus72] Dreyfus L.D., *Intelligence artificielle : Mythes et limites*, Paris : Flammarion, 1984. (Note : Édition originale en anglais publiée en 1972, revue et complétée en 1979)

[Amandry77] Amandry P., Avant-propos [en ligne], In : *Bulletin de Correspondance Hellénique*, *n°101,* Athènes : Ecole française d'Athènes, 1977. p.1-3. Disponible sur Internet : <http://cefael/horde/rayge/detail.php?site_id=1&actionID=page&series_id=BCH&volume_number=101&issue_number=1&page_type=1&page_number=1> (consulté le 26/09/03)

[Andrews96] Andrews K., *Browsing, building, and beholding cyberspace: New approaches to the navigation, construction, and visualisation of hypermedia on the Internet* [en ligne], Technical Sciences Doctor dissertation, Graz University of Technology, 1996. Disponible sur Internet : <http://www2.iicm.edu/keith-phd> (consulté le 26/09/03)

[Aristote-300] Aristote, *Organon : I. Catégories ; II. De l'interprétation* (Trad. J. Tricot), Paris : Vrin, 1959, 153 p.

[Aristotle-300] Aristotle, *Categories & De Interpretatione* (Trad. J.L. Ackrill), Oxford : Clarendon Press, 1963, 162 p.

[Arsac84] Arsac J., L'informatique et le Sens, In : H.L. Dreyfus, *Intelligence Artificielle : Mythes et limites*, Paris : Flammarion, 1984.

[Bachelard71] Bachelard G., *Epistémologie : Textes choisis*, Paris : PUF, 1971, 216 p. (Note : Publication posthume d'extraits d'articles et d'ouvrages)

[Bachimont99a] Bachimont B., L'intelligence artificielle comme écriture dynamique : de la raison graphique à la raison computationnelle [en ligne], In : Petitot J. (Ed.), *Au nom du sens*, Paris : Grasset, 1999. p.290-319. Disponible sur Internet : <http://www.utc.fr/costech/v1/docs/intelligence_artificielle.pdf> (consulté le 09/10/03)

[Bachimont99b] Bachimont B., De l'hypertexte à l'hypotexte : les parcours de la mémoire documentaire [en ligne], In : *Technologie, Idéologies, Pratiques (TIP), numéro spécial « Mémoires collectives »*, 1999. Disponible sur Internet : <http://www.utc.fr/costech/v1/docs/texte_hypotexte.pdf> (consulté le 09/10/03)

[Bachimont01] Bachimont B., Modélisation linguistique et modélisation logique des ontologies : l'apport de l'ontologie formelle, In: *Conférence « Ingénierie des Connaissances » [IC'2001], Grenoble, 25-27 juin, 2001.*

[BaezaYatesEtRibeiroNeto99] Baeza-Yates R., Ribeiro-Neto B. (Ed*.), Modern Information Retrieval*, ACM Press and Addison-Wesley, 1999, 513 p.

[BanerjeeEtMittal94] Banerjee S., Mittal V.O., On the Use of Linguistic Ontologies for Accessing and Indexing Distributed Digital Libraries [en ligne], In : *Digital Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, College Station (Texas), June 19-21 1994.* Disponible sur Internet : <http://www.csdl.tamu.edu/csdl/DL94/paper/banerjee.html> (consulté le 29/09/03)

[Beguin96] Béguin D., *Les antiquisants face à l'informatique et aux réseaux, Internet et les chercheurs* [en ligne], Rapport intermédiaire, Paris : Ecole Normale Supérieure, Département de Sciences Sociales, Novembre 1996. Disponible sur Internet : <http://elias.ens.fr/atelier/articles/ArticleInternetnov96.html> (consulté le 29/09/03)

[BeimelEtAl98] Beimel A., Geller F., Kushilevitz E., The query complexity of finding local minima in the lattice [en ligne], In : *Proceedings of the eleventh annual ACM conference on Computational Learning Theory COLT'98*, 1998. p.294-302. Disponible sur Internet : <http://doi.acm.org/10.1145/279943.280000> (consulté le 29/09/03)

[Benel98] Bénel A., *La Chronique des fouilles : de la bibliothèque à l'Internet* [en ligne], Rapport de stage, EFA, 1998. Disponible sur Internet : <http://www.efa.gr/Informatique/Benel1998.pdf> (consulté le 29/09/03)

[BenelEtAl99] Bénel A., Calabretto S., Pinon J.-M., Indexation "sémantique" de documents archéologiques, In : *Actes du deuxième colloque du chapitre français de l'ISKO, "L'indexation à l'heure d'Internet", Lyon, 21-22 Octobre 1999.* Editions ISKO-France, 2001. p.145-152.

[BenelEtAl00a] Bénel A., Calabretto S., Pinon J.-M., Iacovella A., Vers un outil documentaire unifié pour les chercheurs en archéologie, In : *Actes du XVIIIe congrès INFORSID, Lyon, 16-19 Mai 2000,* Éditions INFORSID. p.133-145.

[BenelEtAl00b] Bénel A., Calabretto S., Pinon J.-M., Iacovella A., Consultation de documents et sémantique : Application à des publications savantes, In : *Actes du second Colloque International Francophone sur l'Ecrit et le Document [CIFED'2000], Lyon, 3-5 Juillet 2000,* Lausanne : PPUR. p.271-280.

[BenelEtCalabretto00] Bénel A., Calabretto S., Exploration de corpus de documents archéologiques à l'aide de théories algébriques. In : *Escuela interlatina de altos*

estudios en linguistica aplicada, Segundo seminario, "Matematicas y tratamiento de corpus", San Millan de la Cogolla, La Rioja, Espana, 19-23 Septiembre 2000, Logroño : Fundacion San Millan de la Cogolla, 2002. p.343-350.

[Benel00] Bénel A., Sémantique interprétative et construction des savoirs scientifiques, In : *Table ronde "Renouvellements méthodologiques dans les bibliothèques numériques et les publications scientifiques", Athènes, Grèce, 18-19 novembre 2000.*

[BenelEtAl01a] Bénel A., Egyed-Zsigmond E., Prié Y. Calabretto S., Mille A., Truth in the Digital Library: From Ontological to Hermeneutical Systems [en ligne], In: *Proceedings of the fifth European Conference on Research and Advanced Technology for Digital Libraries [ECDL'2001], Darmstadt, September 4-9, 2001. Lecture Notes in Computer Science #2163.* Berlin : Springer-Verlag. p.366-377. Disponible sur Internet : <http://lisi.insa-lyon.fr/~abenel/informatique/benel_egyed_ECDL_01.pdf> (consulté le 30/09/03)

[BenelEtAl01b] Bénel A., Calabretto S., Iacovella A., Porphyre : un système pour les bibliothèques numériques, In : *Magazine du département informatique de l'INSA, Dossier "Documentique", Décembre 2001.* p.12-14.

[Benel02a] Bénel A., Ontologies... déontologie, In : *Séminaire annuel de l'Institut des Sciences du Document Numérique, Table-ronde sur le "Web sémantique", Eveux, 22-24 mai 2002.*

[Benel02b] Bénel A., Quels outils documentaires pour les Sciences humaines ? Présentation du système Porphyre, In : *Journées d'études sur les "Modèles opératoires de production et de diffusion des collections scientifiques dans les bibliothèques numériques", Lyon, 29-30 mai 2002.*

[Benel02c] Bénel A., Intertextualité et intersubjectivité dans le système Porphyre : Instrumentation du travail des experts, In : *Assises nationales du GDR I3, Groupe de travail « Documents Multimédia », Nancy, 4-6 décembre 2002.*

[BenelEtAl02] Bénel A., Calabretto S., Iacovella A., Pinon J.-M., Porphyry 2001: Semantics for scholarly publications retrieval [en ligne], In: *Proceedings of the thirteenth International Symposium on Methodologies for Intelligent Systems*

[ISMIS'2002], Lyon, June 26-29, 2002. Lecture Notes in Artificial Intelligence #2366. Berlin : Springer-Verlag. p.351-361. Disponible sur Internet : <http://lisi.insa-lyon.fr/~abenel/informatique/benel_ISMIS_02.pdf> (consulté le 30/09/03)

[Benel03] Bénel A., Lecture, interprétation et constitution de corpus en Sciences Humaines : Aspects épistémologiques et computationnels, A paraître dans les *Suppléments au Bulletin de Correspondance Hellénique.*

[Bennett97] Bennett J.C., JISC/NPO Studies on the preservation of electronic materials: A framework of data types and formats, and issues affecting the long term preservation of digital material [en ligne], *British Library Research and Innovation Report #50*, 1997. Disponible sur Internet : <http://www.ukoln.ac.uk/services/papers/bl/jisc-npo50/bennet.html> (consulté le 30/09/03)

[Berleant00] Berleant D., Models for reader interaction systems [en ligne], In: *Proceedings of the Ninth ACM Conference on Information and Knowledge Management*, ACM Press, 2000, p.127-133. Disponible sur Internet : <http://doi.acm.org/10.1145/354756.354810> (consulté le 30/09/03)

[BernersLee89] Berners-Lee T., *Information Management: A Proposal* [en ligne], Internal report, CERN, 1989. Disponible sur Internet : <http://www.w3.org/History/1989/proposal.html> (consulté le 30/09/03)

[BernersLeeEtAl01] Berners-Lee T., Hendler J., Lassila O., The Semantic Web [en ligne], In*: Scientific American, May 2001.* Disponible sur Internet : <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> (consulté le 30/09/03)

[BessiereEtAl97] Bessière C., Euzenat J., Jeansoulin R., Ligozat G., Schwer S., Raisonnement spatial et temporel, In : *Actes des journées PRC-GDR IA,* Hermès, 1997. p.77-88.

[BieberEtWan94] Bieber M., Wan J., Backtracking in a multiple-window hypertext environment [en ligne], In: *ECHT '94: Proceedings of the ACM European Conference on Hypermedia Technology,* New York: ACM Press, 1994. p.158-166. Disponible sur Internet : <http://doi.acm.org/10.1145/192757.192792> (consulté le 30/09/03)

[BieberEtAl97] Bieber M., Vitali F., Ashman H., Balasubramanian V., Oinos-Kukkonen H., Fourth generation hypermedia: some missing links for the World Wide Web [en ligne], In: *Int. J. Human-Computer Studies #47,* Academic Press Limited, 1997. p.31-65. Disponible sur Internet : <http://ijhcs.open.ac.uk/bieber/bieber.pdf> (consulté le 30/09/03)

[BienzEtAl96] Bienz T., Cohn R., Meehan J.R., *Portable Document Format Reference Manual: Version 1.2*, Adobe Systems Incorporated, 1996.

[Boley98] Boley H., *Declarative Operations on Nets* [en ligne], revised version, DFKI research report, 1998. Disponible sur Internet : <http://www.dfki.uni-kl.de/~boley/drlhops.ps> (consulté le 30/09/03)

[Borillo84] Borillo M., *Informatique pour les Sciences de l'Homme : Limites de la formalisation du raisonnement*, Bruxelles : Pierre Mardaga Éditeur, 1984, 210 p.

[BouaudEtAl94] Bouaud J., Bachimont B., Charlet J., Zweigenbaum P., Acquisition and structuring of an ontology within conceptual graphs [en ligne], In: *Proceedings of ICCS'94 Workshop on Knowledge Acquisition using Conceptual Graph Theory,* 1994. p.1-25. Disponible sur Internet : <http://www.atala.org/~pz/FTPapiers/Bouaud:ICCSW94.ps.gz> (consulté le 30/09/03)

[BrasEtAl90] Bras M., Coulon D., Desclés J.-P., Fuchs C., Gayral F., Jayez J., Kayser D., Nef F., Reppert D., Saint Dizier P., Tollu c., Toussaint Y., Victorri B., La sémantique des langues naturelles : Éléments d'une approche comparative, In *:* *Actes du PRC-GDR IA,* Hermès, 1990. p.369-410.

[Bruneau76] Bruneau Ph., Quatre propos sur l'archéologie nouvelle [en ligne], In : *Bulletin de Correspondance Hellénique*, *n°100,* Athènes : Ecole française d'Athènes, 1976. p.103-130. Disponible sur Internet : <http://cefael.efa.gr/horde/rayge/detail.php?site_id=1&actionID=page&series_id=BCH&volume_number=100&issue_number=1&startpos=105> (consulté le 30/09/03)

[Bruneau92] Bruneau Ph., Le statut archéologique de la trace, In : *Revue d'Archéologie Moderne et d'Archéologie Générale [RAMAGE], n°10,* Paris : Presses Universitaires de l'Université de Paris-Sorbonne, 1992. p.87-93.

[BuffereauEtPicouet02] Buffereau B., Picouet P., Des itinéraires pour accéder à l'information : Un défi pour la représentation de connaissances, In : *Actes de la*

conférence scientifique "Documents virtuels personnalisables" DVP'2002, Brest, Juillet 2002. p.105-114.

[BurrowEtEklund94] Burrow A., Eklund P.W., Visual structure representations and conceptual graphs [en ligne], In: *Proceedings of the fourth international workshop on Peirce: A conceptual graph workbench, Maryland, August 19, 1994.* p.4-10. Disponible sur Internet : <http://citeseer.nj.nec.com/ellis94proceedings.html> (consulté le 30/09/03)

[Bush45] Bush V., As we may think [en ligne], *The Atlantic monthly, #176, July 1945.* p.101-108. Disponible sur Internet : <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm> (consulté le 30/09/03)

[BusnelEtAl97] Busnel F., Godfroy V., Grolleau F., *Le temps : une approche philosophique*, Paris : Ellipses, 1997, 156 p.

[Calabretto93] Calabretto S., *Contribution à la validation des spécifications algébriques et à l'étude des spécifications algébriques avec contraintes,* Thèse de doctorat en informatique, INSA de Lyon, 1993.

[Calabretto03] Calabretto S., Modèles de représentation de la sémantique des documents, A paraître dans les *Suppléments au Bulletin de Correspondance Hellénique.*

[CalvetEtAl01] Calvet M.-A., Lemaître M., Léonhardt J.-L., *Vocabulaires multilingues*, Rapport de mission, Maison de l'Orient Méditerranéen, CNRS, 2001.

[Caplat02] Caplat G., *Modélisation cognitive et résolution de problèmes*, Lausanne : PPUR, 2002, 199 p.

[Caplat03] Caplat G., Connaissance, sémantique, modèle et formalisme, A paraître dans les *Suppléments au Bulletin de Correspondance Hellénique.*

[Caron92] Caron J., *Précis de psycholinguistique*, Presses Universitaires de France, 1992, 275 p.

[CarpinetoEtRomano94] Carpineto C., Romano G. Dynamically bounding browsable retrieval spaces: an application to Galois lattices, In: *RIAO'94 conference proceedings, « Intelligent Multimedia Information Retrieval Systems and Management »*, 1994. p.533-547.

[CavnarEtGillies94] Cavnar W.B., Gillies A.M., Data Retrieval and the Realities of Document Conversion [en ligne], In: *Digital Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, College Station (Texas), June 19-21 1994.* Disponible sur Internet : <http://www.csdl.tamu.edu/csdl/DL94/position/cavnar.html> (consulté le 30/09/03)

[Chabbat97] Chabbat B., *Modélisation multiparadigme de textes réglementaires* [en ligne], Thèse de doctorat en informatique, INSA de Lyon, 1997. Disponible sur Internet : <http://csidoc.insa-lyon.fr/these/1997/chabbat> (consulté le 30/09/03)

[Chalendar97] Chalendar G.(de), *Abstractions de schémas à partir de situations agrégées* [en ligne], DEA de sciences cognitives, Universités Paris-Sud XI, 1997. Disponible sur Internet : <http://www.limsi.fr/Individu/gael/MemoireDEA> (consulté le 30/09/03)

[Charity94] Charity M. N., Multiple Standards? No problem [en ligne], In: *Digital Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, College Station (Texas), June 19-21 1994*. Disponible sur Internet : <http://www.csdl.tamu.edu/csdl/DL94/position/charity.html> (consulté le 30/09/03)

[CharletEtAl99] Charlet J., Daigne M., Leroux V., Ingénierie des patrimoines informationnels de l'établissement de santé, In: *Document numérique, volume 3, n°3-4, décembre 1999.* p.81-99.

[CharletEtAl96] Charlet J., Bachimont B., Bouaud J., Zweigenbaum P., Ontologie et réutilisabilité : expérience et discussion, In : *N. Aussenac-Gilles, P. Laublet, C. Reynaud (Ed.),* "*Acquisition et ingénierie des connaissances*", Cepaduès Editions, 1996. p.69-87.

[Chiaramella99] Chiaramella Y. Approches et modèles en recherche d'informations. In : *XVII° congrès INFORSID, La Garde, France, 2-4 juin 1999.*

[Chiaramella97] Chiaramella Y., Browsing and querying: two complementary approaches for multimedia information retrieval, In: *Hypertext - Information Retrieval - Multimedia, HIM'97, Dortmund, September 29 - October 2, 1997.*

[Chomsky68] Chomsky N., *Le langage et la pensée*, Paris : Payot, 1969, 145 p. (Note : Édition originale en anglais publiée en 1968)

[Chouraqui72] Chouraqui E., Le système d'exploitation automatique de l'inventaire général des monuments et richesses artistiques de France (Formalisation du langage d'analyse), In : *Banques de données archéologiques, Marseille, 12-14 juin 1972.*

[Cleveland91] Cleveland G., *Electronic Document Delivery: Converging Standards and Technologies* [en ligne], UDT Series on Data Communication Technologies and Standards for Libraries, 1991. Disponible sur Internet : <http://www.ifla.org/VI/5/reports/rep2/rep2.htm> (consulté le 30/09/03)

[CoppinEtBrucker02] Coppin G., Brucker F. Les hyper-ontologies. In : *Actes de la conférence scientifique "Documents virtuels personnalisables" DVP'2002, Brest, Juillet 2002*. p.171-182.

[Courbin82] Courbin P., *Qu'est-ce que l'archéologie ? : Essai sur la nature de la recherche archéologique,* Paris : Payot, 1982, 238 p.

[CoxEtGreenberg2000] Cox D., Greenberg S., Supporting collaborative interpretation in distributed groupware [en ligne], In: *Proceedings on the ACM CSCW'2000 Conference on Computer Supported Cooperative Work, 2000.* p.289-298. Disponible sur Internet : <http://doi.acm.org/10.1145/358916.359000> (consulté le 30/09/03)

[Daumas02] Daumas F., Conservation et consultation de documents numériques au CINES, In : *Journées d'études sur les "Modèles opératoires de production et de diffusion des collections scientifiques dans les bibliothèques numériques", Lyon, 29-30 mai 2002.*

[DavidEtAl95] David C., Giroux L., Bertrand-Gastaldy S., Lanteigne D., Indexing as Problem Solving: a Cognitive Approach to Consistency [en ligne], In: *ACSI 95: Annual Conference of the Canadian Association for Information Science, Edmonton (Alberta), June 7-10 1995.* Disponible sur Internet : <http://www.ualberta.ca/dept/slis/cais/david.htm> (consulté le 30/09/03)

[Demoule72] Demoule J.-P., Projet de bibliographie automatique en préhistoire et protohistoire européenne, In : *Banques de données archéologiques, Marseille, 12-14 juin 1972.*

[DenhiereEtBaudet92] Denhière D., Baudet S., *Lecture, compréhension de texte et science cognitive*, Paris : PUF, 1992, 317 p.

[DenoueEtVignollet00] Denoue L., Vignollet L., An annotation tool for web browsers and its applications to information retrieval, In : *RIAO'2000 Conference Proceedings,* "*Content-based multimedia information access*", CID-CASIS, 2000. p.180-195.

[DerrienPeden90] Derrien-Péden D., *Analyse des structures de documents : une approche objet,* Thèse de doctorat, Université de Rennes 1, 1990.

[DesfargesEtHelly91] Desfarges P., Helly B. L'archéologie, système d'information scientifique. In : *Aplicaciones Informaticás en Arqueologia : Teorias y sistemas. Saint-Germain-en-Laye, 1991.*

[Dijkstra84-87] Dijkstra E.W., *Science fiction and science reality in computing* [en ligne], EWD952, 1984-1987. Disponible sur Internet : <http://www.cs.utexas.edu/users/EWD/ewd09xx/EWD952.PDF> (consulté le 30/09/03)

[Dreyfus72] Dreyfus L.D., *Intelligence artificielle : Mythes et limites*, Paris : Flammarion, 1984, 443 p. (Note : Edition originale en anglais publiée en 1972, revue et complétée en 1979)

[Dubois91] Dubois D., *Sémantique et cognition : Catégorie, prototypes, typicalité,* Éditions du CNRS, 1991, 342 p.

[Dubois99] Dubois F., *Archéologie et mode de formation de la nécropole : une approche cognitive*, Rapport de projet de fin d'études, EFA, 1999. 8 p. + Annexes.

[Dubucs96] Dubucs J., La logique depuis Russell, In : *R. Blanché, La logique et son histoire*, Seconde édition revue et complétée, Paris : Armand Colin/Masson, 1996.

[Ducrot72] Ducrot O., Schaeffer J.-M., *Nouveau dictionnaire encyclopédique des sciences du langage*, s.v. "Situation de discours", Paris : Seuil, 1995. p.764-775. (Note : Édition originale en 1972)

[Eco62] Eco U., *L'œuvre ouverte,* Paris : Seuil, 1965, 314 p. (Note : Édition originale en italien publiée en 1962)

[Eco73] Eco U., *Le Signe*, Bruxelles : Labor, 1988, 277 p. (Note : Édition originale en italien publiée en 1973)

[EgyedZsigmondEtAl00] Egyed-Zsigmond E., Prié Y., Mille A., Pinon J.-M., A graph based audio-visual document annotation and browsing system, In: *RIAO'2000 Conference Proceedings, « Content-based multimedia information access »,* CID-CASIS. p.1381-1389.

[EklundEtCole02] Eklund P., Cole R., Structured Ontology and Information Retrieval for Email Search and Discovery [en ligne], In: *Proceedings of the thirteenth International Symposium on Methodologies for Intelligent Systems [ISMIS'2002], Lyon, June 26-29, 2002, Lecture Notes in Artificial Intelligence #2366,* Berlin : Springer-Verlag, p.75-84. Disponible sur Internet : <http://link.springer.de/link/service/series/0558/papers/2366/23660075.pdf> (consulté le 30/09/03)

[EtienneEtAl91] Etienne R., Auda Y., Iacovella A. Spécificité des problèmes d'analyse des données en archéologie : Application à l'analyse des nécropoles. In : *Aplicaciones Informaticás en Arqueologia : Teorias y sistemas. Saint-Germain-en-Laye, 1991.*

[Euzenat97] Euzenat J., A Protocol for Building Consensual and Consistent Repositories [en ligne], Research report, INRIA Rhônes-Alpes, 1997. 46 p. Disponible sur Internet : <http://www.inria.fr/rrrt/rr-3260.html> (consulté le 30/09/03)

[Fargues90] Fargues J., Remarks on the Interrelation between Artificial Intelligence, Mathematical Logic and Humanities, In: *Interpretations in the humanities: Perspectives from Artificial Intelligence*, British Library Board, 1990.

[Farouki96] Farouki N., *La foi et la raison : Histoire d'un malentendu*, Paris : Flammarion, 1996, 321 p.

[Feyerabend61] Feyerabend P.K., *Une connaissance sans fondements*, Editions Dianoïa, 1999, 127 p. (Note : Edition originale en anglais publiée en 1961)

[Fourel96] Fourel F., Intégration de la structure du document dans le processus de recherche d'information, In: *XIVe Congrès INFORSID, Bordeaux, Juin 1996.*

[Fourel98] Fourel F., *Modélisation, indexation et recherche de documents structurés*, Thèse de doctorat en informatique, Université Grenoble 1, 1998.

[Fowler92] Fowler R.H., Wilson B.A., Fowler W.A.L., *Information navigator: An information system using associative networks for display and retrieval* [en ligne], University of Texas - Pan American, Technical report, 1992. Disponible

sur Internet : <http://bahia.cs.panam.edu/info_vis/inf_nav/info_nav_tr_92.html> (consulté le 30/09/03)

[Frauenfelder01] Frauenfelder M., A Smarter Web [en ligne], *Technology Review, November 2001.* Disponible sur Internet : <http://www.ontoprise.de/download/A_Smarter_Web.pdf> (consulté le 30/09/03)

[FutrelleEtZhang94] Futrelle R.P., Zhang X., Large-scale Persistent Object Systems for Corpus Linguistics and Information Retrieval [en ligne], In: *Digital Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, College Station (Texas), June 19-21 1994*. Disponible sur Internet : <http://www.csdl.tamu.edu/csdl/DL94/paper/futrelle.html> (consulté le 30/09/03)

[Gallay86] Gallay A., *L'archéologie demain,* Belfond, 1986.

[GammaEtAl94] Gamma E., Helm R., Johnson R., Vlissides J., *Design Patterns: Elements of Reusable Object-Oriented Software*, Boston, Addison-Wesley, 1994, 395 p.

[Gardin84] Gardin J.-C., Les bases de données dans les sciences de l'antiquité : l'ajustement nécessaire des fins aux moyens, In: *Banques de données et sciences de l'antiquité*, 1984.

[Gardin86] Gardin J.-C., *Systèmes experts et publications savantes, The Fifth British Library Annual Research Lecture,* London : The British Library, 1986, 16 p.

[Gardin96a] Gardin J.-C., La révolution cognitive et l'archéologie. In: *Archeologia e Calculatori, 7, 1996*. p.1221-1230.

[Gardin96b] Gardin J.-C., Formalisation et simulation des raisonnements. In : *Une école pour les sciences sociales*, Paris : Cerf et EHESS, 1996. p.185-208.

[GayteEtAl97] Gayte O., Libourel T., Cheylan J.P., Lardon S.*, Conception des systèmes d'information sur l'environnement*, Hermès, 1997.

[GenestEtChein97] Genest D., Chein M., An Experiment in Document Retrieval Using Conceptual Graphs, In: *Proceedings of the fifth International Conference on Conceptual Structures, Seattle, WA, USA, August 3-8, 1997, LNAI #1257*, Berlin : Springer-Verlag, 1997. p.489-504.

[GenestEtSalvat98] Genest D., Salvat E., A Platform Allowing Typed Nested Graphs: How CoGITo Became CoGITaNT, In: *Proceedings of the sixth International*

Conference on Conceptual Structures: Theory, tools and applications ICCS'98, Montpellier, France, August 10-12, 1998, LNAI #1453, Berlin : Springer-Verlag, 1998. p.154-164.

[Genest99] Genest D., Vers un système de recherche documentaire basé sur les graphes conceptuels, In: *Actes du XVIIe congrès INFORSID, La Garde, France, 2-4 juin 1999*. p.115-131.

[Genest00] Genest D.*, Extension du modèle des graphes conceptuels pour la recherche d'informations*, Thèse de doctorat en Informatique, Université Montpellier II, 2000, 181 p.

[GerstlEtPribbenow95] Gerstl P., Pribbenow S., Midwinters, end games, and body parts: a classification of part-whole relations [en ligne], In: *Formal Ontology in the Information Technology, Special Issue of the International Journal on Human-Computer Studies, vol.43, #5/6, 1995.* Disponible sur Internet : <http://dx.doi.org/10.1006/ijhc.1995.1079> (consulté le 30/09/03)

[GinouvesEtGuimierSorbets78] Ginouvès R., Guimier-Sorbets A.-M., *La constitution des données en archéologie classique*, CNRS Editions, 1978, 161 p.

[GinouvesEtAl98] Ginouvès R. *et al.*, *Dictionnaire méthodique de l'architecture grecque et romaine : III. Espaces architecturaux, bâtiments et ensembles*, s.v. "Thermes", Athènes : Ecole française d'Athènes ; Rome : Ecole française de Rome, 1998. 357 p. – 115 p. de planches.

[Girard89] Girard J.-Y., Le champ du signe ou la faillite du réductionnisme, In : *E. Nagel, J.R. Newman, K. Gödel, J.-Y. Girard, Le théorème de Gödel* (Traduction française et commentaires), Paris : Seuil, 1989.

[Girard95] Girard J.-Y., Intelligence artificielle et logique naturelle, In : *J.-Y. Girard, A. Turing, La machine de Turing* (Traduction française et commentaires), Paris : Seuil, 1995.

[GohEtLeggett00] Goh D., Leggett J., Patron-augmented digital libraries [en ligne], In: *Proceedings of the Fifth ACM Conference on Digital Libraries,* ACM Press, 2000. p.153-163. Disponible sur Internet : <http://doi.acm.org/10.1145/336597.336656> (consulté le 30/09/03)

[Gorse03] Gorse N., Détection d'incohérences dans des spécifications formelles : Application à la modélisation spatiale et temporelle en archéologie, A paraître dans les *Suppléments au Bulletin de Correspondance Hellénique.*

[Grezsiak02] Grezsiak I, *Représentation du Temps et de l'Espace dans les bibliothèques*, Synthèse bibliographique de DESS, ENSSIB, 2002.

[Gruber93] Gruber T.R., A translation approach to portable ontology specifications [en ligne], In: *Knowledge Acquisition, Volume 5, Issue 2, 1993.* p.199-220. Disponible sur Internet : <ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-92-71.ps> (consulté le 30/09/03)

[Guarino97] Guarino N., Some organizing principles for a unified top-level ontology [en ligne], In: *Proceedings of AAAI 1997 Spring Symposium on Ontological Engineering*, AAAI Press, 1997. Disponible sur Internet : <http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/TopLevel.pdf> (consulté le 30/09/03)

[Guarino99] Guarino N., Masolo C., Vetere G., OntoSeek: Content-Based Access to the Web [en ligne], In: *IEEE Intelligent Systems, 14(3), 1999.* p.70-80. Disponible sur Internet : <http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/OntoSeek.pdf> (consulté le 30/09/03)

[Guimier91] Guimier-Sorbets A.-M., Jockey P., Systèmes d'informations sur les sculptures de Délos, In : *Aplicaciones Informaticás en Arqueologia : Teorias y sistemas. Saint-Germain-en-Laye, 1991.*

[Guinaldo95] Guinaldo O., Etude d'un système de gestion de graphes conceptuels, In: *Première journée de l'Ecole Doctorale Sciences pour l'Ingénieur de l'Université Montpellier II, 31 octobre 1995.*

[Hashim90] Hashim S.H., *Exploring hypertext programming: Writing knowledge representation and problem-solving programs*, Windcrest Books Editions, 1990, 365 p.

[Hearst99] Hearst M., User interfaces and visualization [en ligne], In: *R. Baeza-Yates, B. Ribeiro-Neto (Ed.), Modern Information Retrieval,* ACM Press and Addison-Wesley, 1999. p.257-324. Disponible sur Internet : <http://www.sims.berkeley.edu/~hearst/irbook/print/chap10.pdf > (consulté le 30/09/03)

[Helic01] Helic D., *Aspects of Semantic Data Modeling in Hypermedia Systems* [en ligne], Technical Sciences Doctor dissertation, Graz University of Technology, 2001. Disponible sur Internet : <http://www.iicm.edu/thesis/dhelic.pdf> (consulté le 09/10/03)

[Helly02] Helly B., Les expérimentations sur l'image numérique et les bibliothèques virtuelles à la Maison de l'Orient Méditerranéen, In : *Journées d'études sur les "Modèles opératoires de production et de diffusion des collections scientifiques dans les bibliothèques numériques", Lyon, 29-30 mai 2002.*

[HernandezEtAl02] Hernandez A., Ruard-Dumaine F., Desfarges P., La robotisation au cœur d'une chaîne de production, In : *Journées d'études sur les "Modèles opératoires de production et de diffusion des collections scientifiques dans les bibliothèques numériques", Lyon, 29-30 mai 2002.*

[HetzlerEtMiller98] Hetzler B., Miller N., Four critical elements for designing information exploration systems [en ligne], In: *CHI'98 Workshop: Innovation and Evaluation in Information Exploration Interfaces, Los Angeles, California, April 18-23, 1998.* Disponible sur Internet : <http://www.pnl.gov/infoviz/sigchi98> (consulté le 09/10/03)

[Iacovella97] Iacovella A., Etudes des proximités dans l'espace funéraire : Le cas de la nécropole occidentale de Mégara Hyblaea, *Archeologia e Calculatori, 8, 1997*. p.67-102.

[Iacovella02a] Iacovella A., Collections scientifiques et nouvelles technologies : enjeux, méthodes et état des projets, In : *Journées d'études sur les "Modèles opératoires de production et de diffusion des collections scientifiques dans les bibliothèques numériques", Lyon, 29-30 mai 2002.*

[Iacovella02b] Iacovella A., Modèle opératoire de navigation pour les experts : Appropriation sémantique et délimitation de l'espace documentaire, In : *F. Ghitalla (Ed.), La navigation (Numéro spécial), Les cahiers du Numérique, vol.3, n°3,* Paris : Hermès, 2002.

[Iacovella03] Iacovella A., Choix méthodologiques concernant la formalisation des contenus sémantiques dans la documentation numérique : Perspectives scientifiques et expérimentations dans les études archéologiques, A paraître dans les *Suppléments au Bulletin de Correspondance Hellénique.*

[IacovellaEtAl03] Iacovella A., Bénel A. *et al.*, Du partage de corpus documentaires structurés à la confrontation de points de vue, Dossier d'identification d'une équipe projet CNRS STIC, Juillet 2003.

[IFLA96a] *Understanding the UNIMARC format* [en ligne], IFLA Universal Bibliographic Control and International MARC Core Programme, 1996. Disponible sur Internet : <http://www.ifla.org/VI/3/p1996-1/unimarc.htm> (consulté le 03/10/03)

[IFLA96b] *Multi-level Description : Encoding Options for UNIMARC* [en ligne], IFLA Universal Bibliographic Control and International MARC Core Programme. Disponible sur Internet : <http://www.ifla.org/VI/3/p1996-1/guid5.htm> (consulté le 03/10/03)

[IFLA96c] *Guidelines for Using UNIMARC for Component Parts* [en ligne], IFLA Universal Bibliographic Control and International MARC Core Programme. Disponible sur Internet : <http://www.ifla.org/VI/3/p1996-1/guid1.htm> (consulté le 09/10/03)

[IorioEtTurner99] Iorio P. (d'), Turner W.A., Nietzsche sur Internet : L'observation des collaborations médiatisées par ordinateur dans les sciences de l'érudition [en ligne], In : *J. Link Pezet, Les Collaborations Médiatisées par Ordinateurs, Solaris, Dossier nº 5, janvier 1999.* Disponible sur Internet : <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d05/5turner.html> (consulté le 09/10/03)

[Iorio00] Iorio P. (d') (Ed.), *Hypernietzsche : Modèle d'un hypertexte savant sur Internet pour la recherche en sciences humaines. Questions philosophiques, problèmes juridiques, outils informatiques*, Paris : PUF, 2000, 216 p.

[Jacob02] Jacob C., *La lecture assistée par ordinateur* [en ligne], Note de recherche, CNRS Centre Louis Guernet, 2002. Disponible sur Internet : <http://dyonisos.ehess.fr/centres/GDR/Textes%20en%20ligne/Lecture%20savante%20%20Texte.pdf> (consulté le 09/10/03)

[Jockey99] Jockey Ph., *L'archéologie*, Paris : Belin, 1999, 399 p.

[JohnsonEtCochrane95] Johnson E.H., Cochrane P.A., A hypertextual interface for a searcher's thesaurus [en ligne], in: *DL'95 Second Annual Conference on the Theory and Practice of Digital Libraries, Austin, Texas, June 11-13, 1995.*

Disponible sur Internet : <http://www.csdl.tamu.edu/DL95/papers/johncoch/johncoch.html> (consulté le 09/10/03)

[KahanEtAl01] Kahan J., Koivunen M.-R., Prud'Hommeaux E., Swick R.R., Annotea: An open RDF infrastructure for shared web annotations [en ligne], In: *Proceedings of the tenth international conference on World Wide Web,* ACM Press, 2001. Disponible sur Internet : <http://www10.org/cdrom/papers/488> (consulté le 09/10/03)

[KanellosEtAl00] Kanellos I., Thlivitis T., Léger A., Indexation et recherche d'information imagée par le contenu : catégories de textes et catégories d'images dans le projet SemIndex, In : *Actes du troisième Colloque International sur le Document Électronique CIDE'2000, Lyon, 4-6 juillet 2000,* Paris : Europia Productions. p.171-183.

[KanellosEtAl03] Kanellos I., Thlivitis T., Léger A., Indexation et recherche d'information interprétativement tolérante : Une proposition de sémantique textuelle pour les corpus d'images, A paraître dans *les Suppléments au Bulletin de Correspondance Hellénique*.

[Kappe91] Kappe F., *Aspects of Modern Multi-Media Information System* [en ligne], Technical Sciences Doctor dissertation, Graz University of Technology, 1991. Disponible sur Internet : <ftp://ftp.iicm.tu-graz.ac.at/pub/papers/report308.pdf> (consulté le 09/10/03)

[Kasdorf98] Kasdorf B., SGML and PDF, why we need both, In: *Journal of Electronic Publishing* [en ligne], June 1998, Volume 3, Issue 4. Disponible sur Internet : <http://www.press.umich.edu/jep/03-04/kasdorf.html> (consulté le 09/10/03)

[KheirbeckEtChiaramella95] Kheirbek A., Chiaramella Y., Integrating hypermedia and information retrieval with conceptual graphs formalism, In *: Proceedings of HIM'95: Hypertext-Information Retrieval-Multimedia, 1995.* p.47-60.

[KohonenEtKaski00] Kohonen T., Kaski S., Self organization of a massive document collection. In: *IEEE Transactions on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery, Volume 11, Issue 3, 2000.* p.574-585.

[Komninidis95] Komninidis M., *Les archives de l'EFA et la gestion électronique de documents*, DEA en informatique, INSA de Lyon, 1995. 99 p. + Annexes.

[Kuhn62] Kühn T. S., *La structure des révolutions scientifiques*, Paris : Flammarion, 1983, 284 p. (Note : Édition originale en anglais publiée en 1962, revue et augmentée en 1970)

[LabicheEtAl00] Labiche J., Ogier J.-M., Gardes J., Connaissances actionnables et cycles de vie, In : *Actes du troisième Colloque International sur le Document Électronique CIDE'2000, Lyon, 4-6 juillet 2000,* Paris : Europia Productions. p.217-228.

[Larcher97] Larcher F., *Rapport d'installation du SGBI et étude organisationnelle* [en ligne], Rapport de stage, EFA, 1997. Disponible sur Internet : <http://www.efa.gr/Informatique/Larcher1997.pdf> (consulté le 09/10/03)

[Lehmann94] Lehmann F., CCAT: The current status of the conceptual catalogue (Ontology) group with proposals [en ligne], In: *Proceedings of the fourth international workshop on Peirce 'A Conceptual Graph Workbench'*, 1994. Disponible sur Internet : <http://www.robotwisdom.com/ai/fritz.html> (consulté le 09/10/03)

[Leonhardt99a] Léonhardt J.-L., *Test de Turing : Les ordinateurs peuvent-ils penser ? Polémique*, Note de recherche, Maison de l'Orient Méditerranéen, CNRS, 1999.

[Leonhardt99b] Léonhardt J.-L., *Le champ philosophique est-il limité ? ou La connaissance par révélation est-elle redevable de la philosophie ?*, Note de recherche, Maison de l'Orient Méditerranéen, CNRS, 1999.

[Leonhardt02] Léonhardt J.-L., *Analyse sémiotique d'une polémique en archéologie ou Y a-t-il plusieurs modèles de la raison de l'homme de science ?* [en ligne], Note de recherche, Maison de l'Orient Méditerranéen, CNRS, 2002. Disponible sur Internet : <http://www.mom.fr/theme8/Debat.pdf> (consulté le 09/10/03)

[Lepers95] Lepers J.-M., Hypertexte et Infratexte, In : *Hypertextes et hypermédias : Réalisations, outils et méthodes*, Techniques de l'Information, Paris : Hermès, 1995. p. 287-296.

[Lesk96] Lesk M., *Seven Ages of Information Retrieval* [en ligne]. Ottawa : International Federation of Library Associations and Institutions, Universal

Dataflow and Telecommunications Core Programme, March 1996. Disponible sur Internet : <http://www.ifla.org/VI/5/op/udtop5/udtop5.htm> (consulté le 09/10/03)

[Lethbridge94] Lethbridge T.C., *Practical Techniques for Organizing and Measuring Knowledge* [en ligne], Ph.D. thesis in Computer Science, University of Ottawa, 1994. Disponible sur Internet : <http://www.csi.uottawa.ca/~tcl/thesis.pdf> (consulté le 09/10/03)

[Levy81] Lévy P., *L'intelligence collective : Pour une anthropologie du cyberspace*, Paris : La Découverte, 1997, 246 p. Note : Édition originale publiée en 1981.

[LitvakKingEtGarciaMoll72] Litvak King J., García Moll R., Set Theory Models: an Approach to taxonomic and locational (*sic*) relationships. In: *Models in Archaeology, Methuen, London, 1972.*

[LiuEtAl92] Liu Y.H., Pham H.N., Dubuisson B., Reconnaissance de la structure logique d'un document scientifique, In : *CNED 92 : Colloque National sur l'Ecrit et le Document, Nancy, juillet 1992.*

[Liu00] Liu K., *Semiotics in Information Systems Engineering*, Cambridge: University Press, 2000, 218 p.

[Martin96] Martin P., *Exploitation de graphes conceptuels et de documents structurés et hypertextes pour l'acquisition de connaissances et la recherche d'informations* [en ligne], Thèse en informatique, Université de Nice - Sophia Antipolis, 1996. Disponible sur Internet : <ftp://ftp.inria.fr/INRIA/publication/Theses/TU-0431> (consulté le 09/10/03)

[MatthesEtAl01] Matthes F., Niederée C., Steffens U., C-Merge: A Tool for Policy-Based Merging of Resource Classifications [en ligne], In: *Proceedings of the fifth European Conference on Research and Advanced Technology for Digital Libraries [ECDL'2001], Darmstadt, September 4-9, 2001. Lecture Notes in Computer Science #2163.* Berlin : Springer-Verlag. p.352-365. Disponible sur Internet : <http://www.springerlink.com/openurl.asp?genre=article&issn=0302-9743&volume=2163&spage=352> (consulté le 09/10/03)

[Maurer96] Maurer H. (Ed.), *HyperWave: The Next Generation Web Solution* [en ligne], Addison-Wesley, 1996, 635 p. Disponible sur Internet : <http://www.iicm.edu/hwbook> (consulté le 09/10/03)

[MazhoudEtAl95] Mazhoud O., Pascual E., Virbel J., Représentation et gestion d'annotations, In : *Hypertextes et hypermédias : Réalisations, outils et méthodes*, Techniques de l'Information, Paris : Hermès, 1995. p. 127-138.

[McKinley97] McKinley T., *From Paper to Web* [en ligne], Indianapolis (Indiana) : Adobe Press, 1997. Chapter 12, Advanced Searching Techniques. Disponible sur Internet : <http://imagebiz.com/ptweb_12.pdf> (consulté le 09/10/03)

[MechkourEtAl95] Mechkour M., Berrut C., Chiaramella Y., Using a Conceptual Graph Framework for Image Retrieval, In: *The International Conference on Multi-Media Modeling MMM'95, Nov. 14-17, 1995.* p.127-142.

[Menzies99] Menzies T., Cost benefits of ontologies [en ligne], In: *ACM Magazine on Intelligence : New Visions of AI in Practice. Volume 10, Issue 3, 1999.* p.26-32. Disponible sur Internet : <http://doi.acm.org/10.1145/318964.318969> (consulté le 09/10/03)

[Miege95] Miège B., *La pensée communicationnelle,* Grenoble : PUG, 1995, 120 p.

[MitraEtWiederhold00] Mitra P., Wiederhold G., Kersten M.L., A Graph-Oriented Model for Articulation of Ontology Interdependencies [en ligne], In: *Proceedings of the seventh International Conference on Extending Database Technology EDBT'2000, LNCS 1777,* Berlin : Springer-Verlag, 2000. p. 86-100. Disponible sur Internet : <http://www.springerlink.com/link.asp?id=mdebqfhphhtk51l1> (consulté le 09/10/03)

[Montocchio97] Montocchio M., *Comment accéder à des bases de données par le Web ?* [en ligne], Rapport de stage, EFA, 1997. Disponible sur Internet : <http://www.efa.gr/Informatique/Montocchio1997.pdf> (consulté le 09/10/03)

[Mounin68] Mounin G., *Clefs pour la linguistique*, Paris : Seghers, 1968, 172 p.

[Mounin72] Mounin G., *La sémantique*, Seconde édition, Paris : Payot, 1997, 268 p. (Note : Édition originale publiée en 1972, revue et corrigée en 1997)

[Mugnier93] Mugnier M.-L., *On Generalization/Specialization for Conceptual Graphs*, Research report, LIRMM, 1993.

[MugnierEtChein96] Mugnier M-L., Chein M., Représenter des connaissances et raisonner avec des graphes, In: *R.I.A. vol.10, n°1, 1996.* p.7-56.

[Muhr97] Muhr T., *Atlas/ti: Short User's Manual* [en ligne], Berlin : Scientific Software Development, 1997. Disponible sur Internet : <http://www.atlasti.de/manshort.pdf> (consulté le 09/10/03)

[NanardEtAl96] Nanard M., Nanard J., Chauche J., Massotte A.-M., Joubert A., Betaille H., La métaphore du généraliste : Acquisition et utilisation de connaissances macroscopiques sur une base de documents techniques, In : *N. Aussenac-Gilles, P. Laublet, C. Reynaud (Ed.),* "*Acquisition et ingénierie des connaissances*", Cepaduès Editions, 1996.

[NanardEtNanard01] Nanard M., Nanard J., Cumulating and sharing end users knowledge to improve video indexing in a video digital library [en ligne], In: *Proceedings of the first ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'01)*, ACM Press, 2001. p.282-289. Disponible sur Internet : <http://doi.acm.org/10.1145/379437.379683> (consulté le 09/10/03)

[Nelson97] Nelson T.H., Embedded Markup Considered Harmful [en ligne], In: *XML: Principles, Tools, and Techniques, World Wide Web Journal 2(4), Fall 1997.* Disponible sur Internet : <http://www.xml.com/pub/a/w3j/s3.nelson.html> (consulté le 09/10/03)

[Nelson99] Nelson T.H., Xanalogical Structure Needed Now More Than Ever [en ligne], In: *ACM Computing Surveys, Volume 31, Issue 4, 1999*, ACM Press, 1999. Article 33. Disponible sur Internet : <http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/60.html> (consulté le 09/10/03)

[Nelson02] Nelson T.H., *I don't buy in* [en ligne], 2002. Disponible sur Internet : <http://ted.hyperland.com/buyin.txt> (consulté le 02/12/02)

[Neumann58] Neumann J. (von), *L'ordinateur et le cerveau*, Paris : Flammarion, 1996, 129 p. (Note : Édition originale en anglais publiée de manière posthume en 1958)

[Nicolescu96] Nicolescu B., *La Transdisciplinarité : Manifeste*, Monaco : Editions du Rocher, 1996. Extrait disponible sur Internet : <http://perso.club-internet.fr/nicol/ciret/vision.htm> (consulté le 09/10/03)

[NigayEtVernier98] Nigay L., Vernier F., Navigational interaction techniques in the search results space, In: *CHI'98 Workshop: Innovation and Evaluation in Information Exploration Interfaces, Los Angeles, California, April 18-23, 1998.*

[NuxEtBenel02] Nux M., Bénel A., *Quelle licence pour Porphyre ? : Synthèse, Proposition*, Rapport interne, EFA, 2002.

[NuxEtAl02] Nux M., Martin O., Huynh R., Bénel A., *Porphyre 2002 : Manuel d'utilisation du client*, EFA, 2002.

[OHaraEtAl98] O'Hara K., Smith F., Newman W., Sellen A., Student readers' use of library documents: implications for library technologies [en ligne], In: *Conference Proceedings on Human Factors in Computing Systems CHI'98*, New York : ACM Press, 1998. p.233-240. Disponible sur Internet : <http://doi.acm.org/10.1145/274644.274678> (consulté le 09/10/03)

[OrsiEtCavallari1892] Orsi P., Cavallari F.S., Megara Hyblaea, In : *Monumenti Antichi dei Linci, 1,* 1892. Colonnes 799-818 (extrait).

[OunisEtPasca98] Ounis I., Pasca M., RELIEF: Combining Expressiveness and Rapidity into a Single System [en ligne], In: *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* New York : ACM Press, 1998. p. 266-274. Disponible sur Internet : <http://doi.acm.org/10.1145/290941.291007> (consulté le 09/10/03)

[Paepcke96] Paepcke A., Digital libraries: Searching is not enough. What we learned on-site [en ligne]*, D-Lib Magazine, May 1996.* Disponible sur Internet : <http://www.dlib.org/dlib/may96/stanford/05paepcke.html> (consulté le 09/10/03)

[PalowitchEtStewart95] Palowitch C., Stewart D., Automating the Structural Markup Process in the Conversion of Print Documents to Electronic Texts [en ligne], In: *Second Annual Conference on the Theory and Practice of Digital Libraries, Austin (Texas), June 11-13 1995.* Disponible sur Internet : <http://www.csdl.tamu.edu/csdl/DL95/papers/palowitc/palowitc.html> (consulté le 09/10/03)

[Paradis96] Paradis F., *Un modèle d'indexation pour les documents textuels structurés*, Thèse de doctorat en informatique, Université Grenoble 1, 1996.

[Pedersen93] Pedersen G.S., A browser for bibliographic information retrieval, based on an application of lattice theory [en ligne], In: *Proceedings of the Sixteenth*

Annual ACM Conference on Research and Development in Information Retrieval (SIGIR'93), 1993. p.270-279. Disponible sur Internet : <http://doi.acm.org/10.1145/160688.160737> (consulté le 09/10/03)

[PettengillEtArango] Pettengill R., Arango G., Four lessons learned from managing World Wide Web digital libraries [en ligne], In: *DL'95 Second Annual Conference on the Theory and Practice of Digital Libraries, Austin, Texas, June 11-13, 1995.* Disponible sur Internet : <http://www.csdl.tamu.edu/DL95/papers/pettengill/pettengill.html> (consulté le 09/10/03)

[Philippe02] Philippe P., Description de la chaîne de numérisation de la Maison de l'Orient Méditerranéen - Jean Pouilloux, In : *Journées d'études sur les "Modèles opératoires de production et de diffusion des collections scientifiques dans les bibliothèques numériques", Lyon, 29-30 mai 2002.*

[Pinon90] Pinon J.-M., Laurini R., *La documentation multimédia dans les organisations*, Paris : Hermès, 1990, 61 p.

[Pinon97] Pinon J.-M., Calabretto S., Poullet L., Document semantic model: An experiment with patient medical record, In: *Proceedings of ICCC/IFIP Conference on Electronic Publishing [EP' 97]*, ICCC Press, 1997.

[Pitrat84] Pitrat J., Quelques remarques sur "Intelligence artificielle, mythes et limites", In: *H.L. Dreyfus, Intelligence Artificielle : Mythes et limites* (Traduction française et commentaires), Paris : Flammarion, 1984.

[Popper35] Popper K. R., *La logique de la découverte scientifique*, Paris : Payot, 1973, 480 p. (Note : Édition originale en allemand publiée en 1935, puis nouvelle édition en anglais publiée en 1959 et complétée en 1960, 1962, 1965 et 1968)

[Popper72] Popper K.R., *La connaissance objective*, Paris : Flammarion, 1991, 578 p. (Note : Édition originale en anglais publiée en 1972)

[PowellEtFrench98] Powell A.L., French J.C., Using multiple views of a document collection in information exploration, In: *CHI'98 Workshop: Innovation and Evaluation in Information Exploration Interfaces, Los Angeles, California, April 18-23, 1998.*

[Poullet97] Poullet L., Formaliser la sémantique des documents – Un modèle unificateur, In : *Actes du congrès INFORSID*, *Toulouse, 10-13 Juin 1997,* Editions INFORSID. p.339-352.

[PoulletEtAl97] Poullet L., Pinon J.-M., Calabretto S., Semantic Structuring of Documents, In: *Proceedings of IEEE Conference on Data Management Systems [BIWIT' 97], Biarritz, July 1997*. p.118-124.

[PredigerEtWille99] Prediger S., Wille R., The Lattice of Concept Graphs of a Relationnaly (*sic*) Scaled Context [en ligne], In: *Seventh International Conference on Conceptual Structures, LNCS #1640*, Berlin : Springer-Verlag, 1999. p.401-414. Disponible sur Internet : <http://wwwbib.mathematik.tu-darmstadt.de/Math-Net/Preprints/Listen/files/2033.ps.gz> (consulté le 09/10/03)

[Prie95] Prié Y., *Contribution à une clarification des rapports entre Sémantique Interprétative et Informatique* [en ligne], DEA en Informatique, Université Rennes 1, 1995. Disponible sur Internet : <http://liris.cnrs.fr/~yprie/dea.html> (consulté le 09/10/03)

[Prie99] Prié Y., *Modélisation de documents audiovisuels en Strates Interconnectées par les annotations pour l'exploitation contextuelle* [en ligne], Thèse de doctorat en Informatique, INSA de Lyon, 1999. Disponible sur Internet : <http://liris.cnrs.fr/~yprie/these.html> (consulté le 09/10/03)

[Rastier91] Rastier F., *Sémantique et recherches cognitives*, Paris : PUF, 2001, 272 p. (Note : Édition originale publiée en 1991)

[Rastier95a] Rastier F., Le terme : entre ontologie et linguistique [en ligne], In : *La banque des mots, n°7, 1995.* p.35-65. Disponible sur Internet : <http://www.revue-texto.net/Inedits/Rastier_Terme.html> (consulté le 09/10/03)

[Rastier95b] Rastier F., Communication ou transmission [en ligne], In : *Césure, n°8, 1995.* p.151-195. Disponible sur Internet : <http://www.revue-texto.net/Inedits/Rastier_Transmission.html> (consulté le 09/10/03)

[Rastier98] Rastier F., Sens et signification, In : *Protée, printemps 1998.* p.7-18.

[Rastier03] Rastier F., De la sémantique à l'interprétation des objets culturels, A paraître dans les *Suppléments au Bulletin de Correspondance Hellénique.*

[RenearEtAl96] Renear A., Mylonas E., Durand D., Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies [en ligne], In: *Journal of*

Research in Humanities Computing, Oxford: Oxford University Press, 1996. Disponible sur Internet : <http://www.stg.brown.edu/resources/stg/monographs/ohco.html> (consulté le 09/10/03)

[Ricoeur69] Ricœur P., *Le conflit des interprétations : Essais d'herméneutique,* Paris : Seuil, 1969, 501 p.

[Ricoeur86] Ricœur P., *Du texte à l'action : Essais d'herméneutique II,* Paris : Seuil, 1986, 452 p.

[Rijsbergen86] Rijsbergen C.J. (van), A new theoretical framework for information retrieval [en ligne], In: *Proceedings of 1986 ACM Conference on Research and Development in Information Retrieval, 1986.* p.194-200. Disponible sur Internet : <http://doi.acm.org/10.1145/253168.253208> (consulté le 09/10/03)

[Roche00] Roche C., Corporate ontologies and concurrent engineering [en ligne], In: *Journal of Materials Processing Technology #107, Elsevier, 2000.* p.187-193. Disponible sur Internet : <http://dx.doi.org/10.1016/S0924-0136(00)00713-5> (consulté le 09/10/03)

[Roche03] Roche C., La construction d'ontologies : quel constat ? , In : *Journées francophones d'Extraction et de Gestion des Connaissances [EGC'2003], Lyon, 22-24 Janvier 2003.*

[RogerEtAl98] Roger D., Lavandier J., Kolmayer E.*, Navigation et interfaces : Cartes conceptuelles et autres outils*, Rapport de synthèse, CERSI, 1998.

[RoscheisenEtAl95] Röscheisen M., Mogensen C., Winograd T., Beyond browsing: Shared comments, soaps, trails, and on-line communities [en ligne], In: *The Third International World Wide Web Conference, "Technology, Tools and Applications", 1995.* Disponible sur Internet : <http://www.igd.fhg.de/archive/1995_www95/proceedings/papers/88/TR/WWW95.html> (consulté le 09/10/03)

[Roussey01] Roussey C., *Une méthode d'indexation sémantique adaptée aux corpus multilingues,* Thèse de doctorat en Informatique, INSA de Lyon, 2001.

[Salton68] Salton G., *Automatic Information Organization and Retrieval*, McGrow-Hill Inc., 1968.

[Schank84] Schank R.C., *The cognitive computer: On language, learning and artificial intelligence*, Reading (Massachusetts) : Addison Wesley, 1984, 268 p.

[SelvatEtChanier97] Selvat T., Chanier T., Traitement automatique pour la représentation graphique de réseaux lexicaux en apprentissage des langues, In: *Bulletin de Linguistique Appliquée et Générale (BULAG), Année 1996-1997.* p.361-371.

[Shannon48] Shannon C.E., A Mathematical Theory of Communication [en ligne], In: *The Bell System Technical Journal, July-October 1948.* Disponible sur Internet : <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf> (consulté le 09/10/03)

[Sonnenwald98] Sonnenwald D.H., An evolving framework for collaborative information exploration [en ligne], In: *CHI'98 Workshop: Innovation and Evaluation in Information Exploration Interfaces, Los Angeles, California, April 18-23, 1998.* Disponible sur Internet : <http://www.fxpal.com/ConferencesWorkshops/CHI98IE/submissions/long/sonnenwald> (consulté le 09/10/03)

[Sowa87] Sowa J.F., Semantic Networks, In: *Shapiro, S.C., Encyclopaedia of Artificial Intelligence,* Wiley, New-York, 1992. (Note: Edition originale publiée en 1987, revue et corrigée en 1992)

[Sowa00] Sowa J.F., *Knowledge Representation: Logical, Philosophical and Computational Foundations*, Pacific Grove: Brooks/Cole, 2000, 594p.

[Sowa01a] Sowa J.F., Signs, Processes, and Language Games: Foundations for Ontology [en ligne], In: *Ninth International Conference on Conceptual Structures,* "*Broadening the Base*"*, Stanford University, California, July 30 - August 3, 2001.* Disponible sur Internet : <http://www.jfsowa.com/pubs/signproc.htm> (consulté le 09/10/03)

[Sowa01b] Sowa J.F., Automating Ontology Development [en ligne], In: *IJCAI Workshop on Knowledge Discovery, August 6, 2001.* Disponible sur Internet : <http://www.jfsowa.com/pubs/autotalk.htm> (consulté le 09/10/03)

[Stengers93] Stengers I., *L'invention des sciences modernes*, Paris : Flammarion, 1995, 211 p. Note : Édition originale en italien publiée en 1993.

[Stiegler00] Stiegler B., Annotation, navigation, édition électronique : Vers une géographie de la connaissance [en ligne], *Ec/arts, n°2, 2000.* Disponible sur Internet : <http://www.ecarts.org/order/_doc.asp?id=84> (consulté le 09/10/03)

[TalonEtTrigano92] Talon B., Trigano P., La définition : un outil pour l'acquisition de concepts nouveaux dans un lexique sémantique [en ligne], In : *Douzièmes journées internationales sur les systèmes experts et leurs applications, Avignon, juin 1992.* Disponible sur Internet : <http://www.hds.utc.fr/~ptrigano/avi92.html> (consulté le 09/10/03)

[TalonEtTrigano90] Talon B., Trigano P., Enrichissement du lexique : Vers une acquisition "humaine" des entrées [en ligne], In : *ERGO IA 90, Biarritz, Septembre 1990.* Disponible sur Internet : <http://www.hds.utc.fr/~ptrigano/talon90.html> (consulté le 09/10/03)

[TanguyEtThlivitis96] Tanguy L., Thlivitis T., PASTEL : Un protocole informatisé d'aide à l'interprétation des textes, In : *Informatique et Langue Naturelle, Nantes, 1996.*

[Tanguy97a] Tanguy L., *Traitement automatique de la langue naturelle et Interprétation : Contribution à l'élaboration d'un modèle informatique de la Sémantique Interprétative* [en ligne], Thèse de Doctorat en Informatique, Université Rennes 1, 1997. Disponible sur Internet : <http://www.univ-tlse2.fr/erss/textes/pagespersos/tanguy/these> (consulté le 09/10/03)

[Tanguy97b] Tanguy L., Computer-aided Language Processing: Using Interpretation to Redefine Man-machine relations, In: *International Cognitive Technology Conference,* 1997.

[TheBrain01] *Personal Brain: User Guide* [en ligne], TheBrain Technologies Corporation, 2001, 99 p. Disponible sur Internet : <http://www.thebrain.com/products/personalbrain/support/Manual/PB_User_Guide.pdf> (consulté le 09/10/03)

[Tibbo93] Tibbo R.H., *Abstracting, information, retrieval and the humanities*, Chicago : American Library Association, 1993.

[TochtermannEtAlders96] Tochtermann K., Alders T., DogitaLS1: A Digital Library System Based on Hyper-G [en ligne], In: *D-Lib Magazine, October 1996.*

Disponible sur Internet : <http://www.dlib.org/dlib/october96/texas/10tochtermann.html> (consulté le 09/10/03)

[Tochtermann96] Tochtermann K., A First Step Toward Communication in Virtual Libraries [en ligne], College Station (Texas) : Texas A&M University, Center for the Study of Digital Libraries, 1996. Disponible sur Internet : <http://www.csdl.tamu.edu/csdl/pubs/klaus/TecRepKlaus.html> (consulté le 09/10/03)

[Tribollet03] Tribollet R., Gestion hypermédia de corpus documentaires hétérogènes, Mémoire d'ingénieur, Conservatoire National des Arts et Métiers, 2003.

[TriggEtWeiser86] Trigg R.H., Weiser M., TEXTNET: A network-based approach to text handling [en ligne], In : *ACM Transactions on Office Information Systems, Vol.4, No. 1, January 1986.* p.1-23. Disponible sur Internet : <http://doi.acm.org/10.1145/5401.5402> (consulté le 09/10/03)

[Trigg88] Trigg R.H., Guided tours and tabletops: tools for communicating in a hypertext environment [en ligne], In : *CSCW'88: Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, New York: ACM Press, 1988. p.216-226. Disponible sur Internet : <http://doi.acm.org/10.1145/62266.62283> (consulté le 09/10/03)

[Turing50] Turing A.M., Computing Machinery and Intelligence [en ligne], In: *Mind, volume 59, number 236, 1950.* p.433-460. Disponible sur Internet : <http://www.loebner.net/Prizef/TuringArticle.html> (consulté le 09/10/03)

[Turner84] Turner R., *Logiques pour l'intelligence artificielle,* Masson, 1984.

[TVML00] *Interface web de TransVision® : Manuel de référence de TVML*, SGBI Entreprise SA, 2000.

[Ubeda97] Ubeda T., *Contrôle de la qualité spatiale des bases de données géographiques* [en ligne], Thèse de doctorat en informatique, INSA de Lyon, 1997. Disponible sur Internet : <http://csidoc.insa-lyon.fr/these/1997/ubeda> (consulté le 09/10/03)

[UML97] UML Notation Guide. OMG, 1997.

[UNICODE00] *The Unicode Standard: Version 3.0* [en ligne], The Unicode Consortium, Addison-Wesley, 2000. Disponible sur Internet : <http://www.unicode.org/unicode/uni2book/u2.html> (consulté le 09/10/03)

[Vatre01] Vatré R., *Porphyre : Modélisation de parcours de lecture*, Rapport de stage, 2001.

[VaudryEtAl02] Vaudry C., Ranwez S., Poulon A., Crampes M., Initiative mixte dans les DVP : De la pertinence à l'adaptation. In : *Actes de la conférence scientifique "Documents virtuels personnalisables" DVP'2002, Brest, Juillet 2002.* p.141-154.

[VeerasamyNavathe95] Veerasamy A, Navathe S., Querying, navigating and visualizing a digital library catalog [en ligne], In: *Second Annual Conference on the Theory and Practice of Digital Libraries, Austin, Texas, June 11-13, 1995.* Disponible sur Internet : <http://www.csdl.tamu.edu/DL95/papers/veerasamy/veerasamy.html> (consulté le 09/10/03)

[Veron97] Veron M., *Modélisation de la composante annotative dans les documents électroniques* [en ligne], Mémoire de DEA RCFR INPT, Toulouse, Septembre 1998. Disponible sur Internet : <http://www.enseeiht.fr/fr/recherche/info/Intellig/PUBLICATIONS/VERON/dea_mv.txt> (consulté le 09/10/03)

[WangBaldonado97] Wang Baldonado M.Q., *An Interactive, Structure-Mediated Approach to Exploring Information in a Heterogeneous, Distributed Environment*, Ph.D. dissertation in Computer Science, Stanford University, 1997.

[WangBaldonadoEtWinograd97] Wang Baldonado M.Q., Winograd T., SenseMaker: An information-exploration interface supporting the contextual evolution of a user's interests [en ligne], In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'97), Atlanta, Georgia, April 1997.* p.11-18. Disponible sur Internet : <http://doi.acm.org/10.1145/258549.258563> (consulté le 09/10/03)

[WeibelEtAl98] Weibel S., Kunze J., Lagoze C., Wolf M., *Dublin Core Metadata for Resource Discovery* [en ligne], The Internet Society, RFC #2413, September 1998. Disponible sur Internet : <ftp://ftp.isi.edu/in-notes/rfc2413.txt> (consulté le 09/10/03)

[WittenEtAl94] Witten I.H., Moffat A., Bell T.C., *Managing Gigabytes: Compressing and Indexing Documents and Images*, Second Edition, San Francisco : Morgan Kaufmann Publishing, 1999, 518 p. (Note : First edition published in 1994)

[Wood02] Wood L., The World in a Box: Little fanfare greets the coming out of a pivotal AI project [en ligne], *Scientific American, January 2002.* Disponible sur Internet : <http://www.sciam.com/print_version.cfm?articleID=00063887-5C1E-1C6D-84A9809EC588EF21> (consulté le 09/10/03)

[ZackladEtFontaine96] Zacklad M., Fontaine D., L'acquisition des connaissances classificatoires pour les systèmes à base de connaissances, In : *N. Aussenac-Gilles, P. Laublet, C. Reynaud (Ed.), Acquisition et ingénierie des connaissances*, Cepaduès Editions, 1996. p.187-206.

[Zaid99] Zaïd N., La publication archéologique sur Internet, *Archaeologia, janvier 1999, n°352.* p.20-29.

[Zramdini92] Zramdini A., Azokly A., Ingold R., Importance de l'identification de la fonte dans la reconnaissance structurelle de documents. In : *CNED'92 : Colloque National sur l'Ecrit et le Document, Nancy, juillet 1992.*

[ZweigenbaumEtAl95] Zweigenbaum P., Bachimont B., Bouaud J., Charlet J., Boisvieux J.-F., Issues in the structuring and acquisition of an ontology for medical language understanding [en ligne], In: *Methods of Information in Medicine, 34 (1/2), 1995.* p.15-24. Disponible sur Internet : <http://www-test.biomath.jussieu.fr/~pz/FTPapiers/Zweigenbaum:MIM95.pdf> (consulté le 09/10/03)

[ZweigenbaumEtAl97] Zweigenbaum P., Bouaud J., Bachimont B., Charlet J., Séroussi B., Boisvieux J.-F., From text to knowledge: a unifying document-centered view of analyzed medical language [en ligne], In: *Fourth International Conference on Medical Concept Representation, Jacksonville, 1997*. Disponible sur Internet : <http://www-test.biomath.jussieu.fr/~pz/FTPapiers/Zweigenbaum:ICMCR97.pdf> (consulté le 09/10/03)

Table des matières

Introduction 1
1ère partie : État de l'art 3
Chapitre 1. Préliminaires méthodologiques 5
1. Une problématique issue des sciences historiques 5
2. Une approche à la fois interdisciplinaire et opératoire 7
a. Mise en ligne des collections scientifiques de l'EFA 7
b. Recherches sur la gestion de la sémantique des collections scientifiques 8
3. Cartographie des écueils 10
a. Création d'une nouvelle discipline 10
b. Instrumentalisation et réification des disciplines 11
c. Impérialisme d'une discipline 11
d. Dissymétrie des « produits » 11
4. Définition de l'objet interdisciplinaire 12
a. Recherche d'informations 13
b. Modélisation des connaissances 14
c. Hypertexte 15
d. Quelle grille de lecture ? 15
Chapitre 2. Statut des modèles informatiques : Le cas des ontologies 17
1. Valeur ontologique 18
a. Définition 18
b. Le mythe de l'intelligence artificielle 19
c. Courant philosophique 21
2. Valeur logique 22
a. Définition 22
b. L'obsession de l'interopérabilité 23
c. Courant philosophique 24
3. Valeur épistémologique 26
a. Définition 26
b. Des machines qui donnent à penser 27

c. Courant philosophique 28
4. Conclusion 30
Chapitre 3. Explication et compréhension en Sciences Humaines 31
1. Le cas de la formalisation en archéologie 31
2. L'indispensable compréhension 33
3. Un cadre interprétatif pour le calcul 36
Chapitre 4. Instrumentation de l'intertextualité et de l'intersubjectivité 39
1. Au cœur des usages 39
2. Panorama des systèmes hypermédia 41
a. Xanadu® 41
b. World Wide Web 42
c. Hyper-G 44
d. ATLAS.ti® 45
e. PASTEL 46
f. Strates-IA 48
g. TheBrain® 49
3. Bilan provisoire 50
2ème partie : Modèle proposé 53
Chapitre 5. Traces et publication 55
1. Notion de trace 56
a. Principe 56
b. Architecture 57
2. Notions de publication, de communauté et d'édition 60
3. Histoire d'une trace 61
Chapitre 6. Contenus documentaires 65
1. Notions 66
a. Contenu documentaire 66
b. Objet documentaire 66
2. Traitements 69
a. Sur un objet documentaire isolé 69
b. Sur un contexte de lecture 70

Chapitre 7. Réseaux de description 73
1. Décrire de manière structurée des objets documentaires 74
a. Des arbres qui cachent… la bibliothèque 74
b. Alternatives 76
c. Réfutation 77
2. Un modèle à base de points de vue 79
3. Arpenter l'espace documentaire 81
a. Filtrer des descripteurs « à plat » 82
b. Filtrer des réseaux de descripteurs 85
c. Scénario de recherche de documents 89
d. Scénario d'indexation de documents 91
4. Bilan 94
Chapitre 8. Parcours de lecture 95
1. Principe 96
a. Historique 96
b. Etape de lecture 97
c. Navigation 98
2. Spécification 98
3. Scénario : Feuilleter un ouvrage 101
4. Originalité du modèle 103
3ème partie : Études de cas 105
Chapitre 9. La *Chronique des fouilles* du BCH 107
1. Etude de l'existant 108
a. Présentation du corpus 108
b. « Chronique » ou « Topique » des fouilles ? 109
2. Expérimentation 110
3. Retour d'expérience 112
a. Limites rencontrées 112
b. Solutions proposées 112
Chapitre 10. CEFAEL : Collections de l'Ecole française d'Athènes en ligne 115
1. Etude des besoins 116
a. Maquettiste 117

b. Bibliothèque 117
c. Photothèque/Planothèque 118
d. Equipe de fouille 119
2. Numérisation et valorisation 120
3. Expérimentation dans Porphyre 121
4. Retour d'expérience 122
Chapitre 11. La nécropole occidentale de Mégara Hyblaea 125
1. « Fouiller » un rapport de fouille 125
2. Retour d'expérience 126
a. Premiers résultats 126
b. Et au-delà ? 127
3. Esquisse d'un « guide des bons usages » 128
a. Modéliser une taxinomie 128
b. Modéliser des attributs et des valeurs d'attribut 129
c. Modéliser une méréonymie 129
d. Modéliser des instanciations 130
e. Modéliser des liens d'association non orientés 131
f. Modéliser des associations non-orientées 131
g. Modéliser des associations orientées 132
h. Utiliser des documents comme « points de repère » 133
Chapitre 12. Perspectives 135
1. Espace intersubjectif 135
a. principe 135
b. Assistance à la pose de contraintes inter- points de vue 136
c. Consensus et importation 136
2. Espace diachronique 137
3. Application au temps archéologique 137
a. Présentation 137
b. « Drôle de temps » 137
c. Travaux actuels et à venir 138
4. Application au document d'architecture (en archéologie) 139
a. Présentation 139

Conclusion 143
Bibliographie 147
Table des matières 177
Annexes 187
Annexe A : Syntaxe utilisée pour les spécifications algébriques I
Annexe B : Historique du prototype III
Porphyre 1999 III
Porphyre 2000 III
Porphyre 2001 III
Porphyre 2002 III
Porphyre 2003 III
Annexe C : Pourquoi diffuser *Porphyre* sous Licence GPL ? V
Annexe D : Manuel d'utilisation du client *Porphyre* VII
FOLIO ADMINISTRATIF X

Liste des figures et tableaux

Figure 2.1 : Valeur ontologique du discours 19
Figure 2.2 : Valeur logique du discours 22
Figure 2.3 : Extrait d'un exemple de sérialisation d'arbre en XML 24
Figure 2.4 : Extrait d'un exemple de sérialisation de graphe en RDF 24
Figure 2.5 : Valeur épistémologique du discours 26
Figure 3.1 : L'interprétation d'après Dilthey comme compréhension pure – disjointe de l'explication (diagramme d'activité UML) 34
Figure 3.2 : L'interprétation d'après Paul Ricœur comme compréhension et explication mises en relation par le texte (diagramme d'activité UML) 35
Figure 3.3 : L'explication par le calcul d'après François Rastier : le passage obligé de la signification au symbole – et inversement (diagramme d'activité UML) 37
Figure 4.1 : Rétro-conception de *Xanadu*® (diagramme de classe UML) 41
Figure 4.2 : Rétro-conception du *World Wide Web* (diagramme de classe UML) 43
Figure 4.3 : Rétro-conception d'*Hyper-G* (diagramme de classe UML) 45
Figure 4.4 : Rétro-conception d'*ATLAS.ti*® (diagramme de classe UML) 46
Figure 4.5: Rétro-conception de *PASTEL* (diagramme de classe UML) 47
Figure 4.6: Rétro-conception de *Strates-IA* (diagramme de classe UML) 48
Figure 4.7 : Rétro-conception de *TheBrain*® (diagramme de classe UML) 50
Figure 5.1 : Notion de trace 57
Figure 5.2 : Types de traces et liens contextuels 57
Figure 5.3 : Architecture multi-tiers du système *Porphyre* pour un accès natif (diagramme de déploiement UML) 58
Figure 5.4 : Architecture multi-tiers du système *Porphyre* pour un accès par le Web (diagramme de déploiement UML) 59
Figure 5.5 : La publication comme moteur de l'interprétation 61
Figure 5.6 : « Cycle de vie » d'un contexte (diagramme d'état UML) 62
Figure 6.1 : Exemple de correspondance pour un contenu documentaire entre une référence automatique et un nom significatif. 67
Figure 6.2 : Contenus documentaires (diagramme de classe UML) 68

Figure 6.3 : A partir d'un même contenu documentaire : a. une vignette, b. une vue de la source, c. une vue d'un fragment. 70
Figure 6.4 : Fonction d'encadrement automatique dans un contexte de lecture comprenant des fragments et leur source. 71
Figure 7.1 : « L'arbre de Porphyre » : Fresque du XVIII[e] s. et diagramme de classe UML. 75
Figure 7.2 : Indexation à l'aide des graphes conceptuels d'une monographie traitant des timbres amphoriques thasiens. 76
Figure 7.3 : Modèle du domaine nécessaire à l'indexation de la figure précédente : concepts, relations et instances. 77
Figure 7.4 : Mosaïque noire sur fond blanc ou blanche sur fond noir ? [Bruneau76] 77
Figure 7.5 : Exemple de réseau de description 79
Figure 7.6 : Réseau de description (diagramme de classe UML) 80
Figure 7.7 : Structure en treillis de l'espace des documents 83
Figure 7.8 : Structure en treillis de l'espace des descripteurs 83
Tableau 7.1 : Exemple de correspondance entre des documents et des descripteurs 84
Tableau 7.2 : Correspondance entre les requêtes et les corpus de documents (calculée à partir de l'exemple du tableau 7.1) 84
Figure 7.9 : Diagramme de classe (notation UML) dérivé de l'espace des documents.. 84
Figure 7.10 : Diagramme d'état (notation UML) dérivé de l'espace des descripteurs ... 85
Figure 7.11 : Mise à jour des filtres dans les facettes « typologie », « espace » et « temps » après changement de sélection dans la facette typologie (diagramme de séquence UML). 88
Figure 7.12 : Requêtes envoyées au serveur lors de la mise à jour des filtres (cf. Figure précédente) 89
Figure 7.13 : Exemple de réseau de description 89
Figure 8.1 : Exemple de navigation entre des étapes de lecture (Diagramme d'état UML). 96
Figure 8.2 : Parcours de lecture (diagramme de classe UML) 99
Figure 8.3 : Exemple de parcours de lecture 102
Tableau 9.1 : Volumétrie de la *Chronique des fouilles* 108
Figure 9.1 : La *« Topique des fouilles»,* virtuellement présente dans la *Chronique* 109

Figure 9.2 : Réseau de description présentant deux facettes pour un même corpus 110
Figure 9.3 : Lecture avec *Porphyre 2000* d'un extrait de la *Chronique des fouilles* 111
Figure 9.4 : Prise en compte du retour d'expérience : nouvelle structuration de la *Chronique des fouilles* à l'aide de Porphyre 113
Figure 10.1 : Une page à étudier sous différents points de vue. 116
Figure 10.2 : Extrait de la facette du maquettiste (Réseau de description *Porphyre*)... 117
Figure 10.3 : Extrait de la facette du bibliothécaire (Réseau de description *Porphyre*)118
Figure 10.4 : Extrait de la facette de l'archiviste (Réseau de description *Porphyre*) ... 119
Figure 10.5 : Extrait de la facette d'une équipe de fouille (Réseau de description *Porphyre*) 120
Figure 10.6 : Lecture avec *Porphyre 2001* d'un extrait des *Collections de l'École française d'Athènes en ligne* 122
Figure 11.1 : Lecture active avec *Porphyre 2002* d'un extrait de la publication des fouilles de la nécropole de Mégara Hyblaea 127
Figure 11.2 : Modélisation d'une taxinomie à l'aide des réseaux de description 128
Figure 11.3 : Taxinomie avec composition : à utiliser avec discernement 129
Figure 11.4 : Modélisation d'attributs à l'aide des réseaux de description..... 129
Figure 11.5 : Méréonymie utilisée pour décrire une chronologie (celle de G. de Mortillet [Jockey99])..... 130
Figure 11.6 : Instanciation multiple 131
Figure 11.7 : Lien d'association « sont au même niveau » 131
Figure 11.8 : Association non-orientée « Tête-bêche » 132
Figure 11.9 : Association orientée « contient »..... 132
Figure 12.1 : Exemple de savoir de référence : structure du lexique et structure du document 139
Figure 12.2 : Exemple d'annotation d'un document d'architecture par un expert 140

Annexes

Annexe A : Syntaxe utilisée pour les spécifications algébriques

Pour les spécifications algébriques développées dans cette thèse nous utilisons :

- les formes de Skolem (cf. *Prolog*),
- les opérations sur les ensembles (intersection, union, différence, définition en extension, définition en intension…),
- les fonctions,
- les formules ayant pour but « $\perp$ » (contradiction formelle) afin d'exprimer des contraintes à vérifier lors de la modification des données.

NOTES :

- L'ensenble *Boolean* et la valeur *NULL* sont prédéfinis.
- Comme en *Prolog,* le caractère de soulignement (« _ ») désigne une variable muette. Dans une même formule, chacune de ses occurrences pourra prendre des valeurs différentes.
- Si E est un ensemble, on considère que $E^1 \neq E$. Par exemple : $\{\{e_0,..e_n\}\} \neq \{e_0,..e_n\}$.
- Afin de s'approcher de la syntaxe « orientée objet », la plupart des fonctions sont notées comme opérateurs post-fixés et ont un nom commençant par un point.

EXEMPLES :

$grandParent(x,z) \leftarrow parent(x,y) \wedge parent(y,z)$

Se lit « Si *x est parent de y* et que *y est parent de z* alors *x est grandParent de z* ».

$\perp \leftarrow parent(x,y) \wedge parent(y,x)$

Se lit « On ne peut avoir à la fois *x parent de y* et *y parent de x* ».

$x.getChild = \{y | parent(x,y)\}$

Se lit « La fonction *getChild* appliquée à *x* renvoie l'ensemble des *y* pour lesquels $parent(x,y)$ est vraie ».

Annexe B : Historique du prototype

Porphyre 1999

Aurélien Bénel (conception générale, réseau de description : noyau)
Franck Eyraud (réseau de description : visualisation)

Porphyre 2000

Aurélien Bénel (conception générale, gestion d'équipe, réseau de description : client)
Thomas Buisson (réseau de description : serveur & base de données)
Mehdi Lababidi (contenus documentaires)

Porphyre 2001

Aurélien Bénel (conception générale, gestion d'équipe, client)
Thomas Buisson (réseau de description : calcul parallèle)
Laurent Pinel (contenus documentaires : servlet)
Elodie Tasia (réseau de description : passerelle Web)
Rodolphe Vatré (parcours de lecture : base de données)
Jocelyn Viallon (réseau de description & contenus documentaires : import)

Porphyre 2002

Aurélien Bénel (conception générale, manuel d'utilisation)
Olivier Chadenat (contenus documentaires : PHP)
Rémi Huynh (parcours de lecture, manuel d'utilisation)
Olivier Martin (notes de lecture, manuel d'utilisation)
Michel Nux (gestion d'équipe, licence « open source », manuel d'utilisation)

Porphyre 2003

Tiphaine Accary (manuel d'installation, site porphyry.org)
Aurélien Bénel (conception générale)
Guillaume Deshors (nouvelle architecture client-serveur)
Caroline Djambian (veille concurrentielle et veille brevet)
Julien Gossa (nouvelle architecture client-serveur)
Baptiste Meurant (nouvelle architecture client-serveur)
Michel Nux (gestion d'équipe, communauté « open source »)
Régine Tribollet (contenus documentaires : PHP)

Annexe C : Pourquoi diffuser *Porphyre* sous Licence GPL ?

Ecole Française d'Athènes – Service informatique
6 rue Didotou, 106 80 Athènes, Grèce.
Téléphone : +30 210 36 79 949

Quelle licence pour Porphyre ?

- Synthèse -

Version 2.1

Dernière modification : 13/08/02

Nombre de pages : 4

Rédacteurs : Aurélien BENEL
Michel NUX

1. Introduction

- **Présentation du projet**

Porphyre s'adresse à des communautés d'experts travaillant avec des documents (chercheurs, ingénieurs, juristes, médecins...). Il vise à stocker une structure hypermédia sur des documents numérisés (pour le moment : texte intégral et image) afin de représenter différents parcours d'interprétation. Il fournit des outils d'aide à la création, à l'organisation, à l'annotation et à la publication de contenus documentaires enrichis de leurs structures hypermédias. D'un point de vue plus technique, les données sont stockées dans des bases de données, traitées par des serveurs distribués et manipulées par les utilisateurs grâce à un client natif ou depuis le Web.

- **Objet du document**

Ce document traite du choix d'une licence pour *Porphyre.* En effet, la licence d'un logiciel est un texte spécifiant les permissions concernant son utilisation, sa copie, sa distribution, sa modification, etc. Le choix d'une licence parmi l'éventail disponible est donc décisif pour l'avenir du logiciel. A ce titre, il doit correspondre aux objectifs à long terme de l'Ecole française d'Athènes.

2. Le prototype Porphyre

- **Situation actuelle**

Le prototype *Porphyre* est le fruit de cinq ans de recherches et de développements initiées par l'Ecole française d'Athènes. Il a été soutenu par le Ministère de la Recherche et le Ministère de l'Education Nationale et de l'Enseignement Supérieur. Il a fait l'objet de collaborations avec le LIRIS (laboratoire de recherche) et le département informatique de l'INSA de Lyon. Il s'est concrétisé au travers d'une thèse et de nombreux stages. Aujourd'hui, il se présente sous la forme d'un prototype opérationnel.

- **Prévisions**

Qu'adviendra-t-il de ce prototype dans un an, deux ans, voire dans plusieurs années ? Le premier danger pesant sur tout logiciel est l'oubli, la plupart des logiciels développés terminent en effet leur vie dans un placard avant même d'avoir été utilisés. Le second danger serait que le projet s'enlise, c'est-à-dire que l'Ecole française d'Athènes soit obligée de le continuer « coûte que coûte » afin de ne pas perdre son investissement en temps et en argent. Le troisième, enfin, serait la récupération du logiciel par une entreprise privée (par des moyens légaux ou non), ce qui se traduirait par une perte de contrôle de la part du secteur public vis-à-vis de l'utilisation et de l'évolution du produit.

C'est pour contrer cela qu'il semble utile de préciser les objectifs qui ont donné le jour au projet et de s'assurer que la licence retenue sera en adéquation.

- **Objectifs**

Les objectifs principaux qui motivent le choix d'une licence sont régis par les points suivants :

✓ *Paternité de l'EFA*

L'Ecole française d'Athènes est à l'origine du projet *Porphyre.* Elle a assuré, grâce à des collaborations interdisciplinaires, les recherches et les développements permettant d'obtenir un prototype opérationnel. Ce projet complexe s'est inscrit dans la durée : de 1998 à 2002. A ce titre, il faut que l'investissement de l'Ecole apparaisse clairement dans le projet (ainsi que dans ses versions futures). En effet, l'Ecole doit pouvoir, si elle le souhaite, rentabiliser cet investissement en terme de prestige (entre autres).

✓ *Un logiciel utilisable par tous*

Fruit d'un financement public, il est capital que *Porphyre* soit utilisable par tous. Par conséquent, il faut assurer sa diffusion publique et le protéger contre toute récupération qui viserait à en faire un logiciel propriétaire.

✓ *Une synergie autour du projet*

Porphyre n'est qu'au début de son évolution que ce soit en terme de développement ou de recherche interdisciplinaire (SHS-STIC : Sciences de l'Homme et de la Société – Sciences et Technologies de l'information et de la Communication). Il est indispensable que toute institution publique (entre autres) puisse, si elle le souhaite, participer à l'évolution du produit. Par conséquent, il faut permettre l'amélioration du logiciel par tous. Plus précisément, il s'agit d'organiser, autour d'un animateur, une communauté de développement coopératif .

3. Différentes solutions

Droit \ Licence	Domaine public	GNU LGPL	GNU GPL	Shareware	Propriétaire
Exécuter	libre	libre	libre	limité (fonctions, données…)	limité (nombre de postes…)
Copier - Distribuer	libre	libre à condition de conserver la licence apposée et les copyrights	libre à condition de conserver la licence apposée et les copyrights	libre	très limité
Etudier	libre	libre	libre	(renforcé par la possibilité de brevets) NON SAUF : pour l'interopérabilité et la recherche	
Intégrer	libre	libre à condition de conserver dans le module réutilisé sa licence et ses copyrights	le produit final doit lui même être sous GPL et faire mention des copyrights originaux	NON	NON
Modifier	libre	libre à condition de conserver la licence apposée et les copyrights	libre à condition de conserver la licence apposée et les copyrights	NON	NON

4. La solution retenue

- **Le choix de la GNU GPL**

Parmi les licences présentées, seule la licence GNU GPL (General Public License) est en totale adéquation avec les objectifs précédemment cités et ceci par une protection inaliénable (portant à la fois sur les évolutions de *Porphyre* et sur les produits dérivés).

En ce qui concerne la documentation liée au projet, la licence GNU FDL (Free Documentation License) peut être choisie afin d'offrir une protection comparable à celle de la GPL pour le logiciel et ses codes sources.

- **Avantages supplémentaires**

Par ailleurs, autour des projets sous licences GNU s'est développée une communauté dynamique et bien organisée dont le but est le développement coopératif.

L'usage veut que ces projets soient gérés par une équipe resserrée (créateurs ou successeurs) afin d'animer et orienter le projet, ainsi qu'assurer l'intégration des modifications.

La communauté offre des services pour la gestion de ce type de projets (sites web, listes de diffusion, forums, gestion de versions, etc.).

Enfin, la communauté dispose de puissants outils de communication. Outre un site web référençant tous les projets, elle dispose de pages réservées dans les magazines Linux du monde entier (éditorial « Brave GNU World »). Il s'agit donc d'un fantastique vecteur permettant de susciter des collaborations parmi la communauté internationale des développeurs et des chercheurs.

5. Conclusion

Par l'application de la licence GNU GPL à l'ensemble des codes sources de *Porphyre* et de la licence GNU FDL à l'ensemble de sa documentation, nous garantissons :

- la reconnaissance de la paternité de l'EFA sur les cinq années passées de recherche et de développement,
- la vocation publique du produit,
- une évolution du produit à laquelle tous ceux qui le souhaitent pourront participer.

Annexe D : Manuel d’utilisation du client *Porphyre*

Ecole Française d'Athènes – Service informatique
6 rue Didotou, 106 80 Athènes, Grèce.
Email : Andrea.Iacovella@efa.gr – Téléphone : +30 210 36 79 949

– Porphyre 2002 –

Manuel d'utilisation du Client

Référence du document :

Pophyre-ManualUtilisationClient-2002fr

Date de création : 09/07/02

Dernière modification : 04/09/02

Etat du document : validé

Nombre de pages : 22

Rédacteurs : Michel NUX
Rémi HUYNH
Olivier MARTIN
Aurélien BENEL

Équipe : Informatique

Validation : Chef de Projet ☑

Responsable Projet ☑

PORPHYRY - Digital workshop for building and handling library collections

LICENSE : GNU FDL

MAINTAINER : Aurelien BENEL (Aurelien.Benel@porphyry.org)

SCIENTIFIC COMMITTEE

- Andrea IACOVELLA (EFA)
- Aurelien BENEL (LIRIS)

OFFICIAL WEB SITE : http://www.porphyry.org

Copyright (C) 2002 Aurelien BENEL, Remi HUYNH, Olivier MARTIN, Michel NUX.
Authors declare by the present not to claim rights on the manual of the program " PORPHYRY " realized for the Ecole Francaise d'Athenes (EFA - Didotou, 6 - 106 80 Athens - Greece - efa@efa.gr).

LEGAL ISSUES
Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1 or any later version published by the Free Software Foundation; with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
A copy of the license is included in the section "legal" entitled "Porphyry_documentation_license.txt".

SOMMAIRE

1. INTRODUCTION 4

1.1. OBJET DU DOCUMENT 4
1.2. TERMINOLOGIE 4

2. LES PREMIERS PAS AVEC PORPHYRE 5

2.1. LES PARAMETRES DE LA CONNEXION 5
2.1.1. pour le serveur Porphyre 5
2.1.2. Pour le serveur FTP 5
2.1.3. Pour le serveur Web 6
2.2. FENETRE DE L'ENVIRONNEMENT DU CLIENT PORPHYRE 7

3. RESEAU DE DESCRIPTION 8

3.1. NAVIGATION 8
3.2. FONCTIONNALITES DU MENU 9
3.2.1. Créer un dossier 9
3.2.2. Ajouter un document 10
3.2.3. Créer un fragment 11
3.2.4. Créer une note 12
3.2.5. Supprimer un descripteur, un document ou une note 13
3.3. FONCTIONNALITES SUPPLEMENTAIRES 14
3.3.1. Déplacer 14
3.3.2. Créer un lien 15
3.3.3. Supprimer un lien 15
3.3.4. Propriétés 15

4. EDITEUR DE NOTES 16

4.1. CRÉATION D'UNE *NOTE* 16
4.2. AJOUT DE TEXTE ET SAUVEGARDE DE LA NOTE 16
4.3. MISE EN FORME DU TEXTE 17
4.4. RETRAIT DE TOUTES LES MISES EN FORME 17
4.5. AJOUT DE CARACTERES SPECIAUX 18

5. PARCOURS DE LECTURE 20

5.1. GENERALITE 20
5.2. LA GESTION DES PARCOURS 20
5.3. LA GESTION DES ETAPES DE LECTURE 21

1. Introduction

Porphyre s'adresse à des communautés d'experts travaillant avec des documents (chercheurs, ingénieurs, avocats, médecins...). Il vise à stocker une structure hypermédia sur des documents numérisés (pour le moment : texte intégral et image) afin de représenter différents parcours d'interprétation. Il fournit des outils d'aide à la création, à l'organisation, à l'annotation et à la publication du contenu de document enrichi de sa structure hypermédia. D'un point de vue plus technique, les données sont stockées dans des bases de données, traitées par des serveurs distribués et manipulées par les utilisateurs grâce à un client natif ou depuis le Web.

1.1. Objet du document

Le but de ce document est d'expliquer l'utilisation des fonctionnalités offertes par *Porphyre* à travers des exemples.

1.2. Terminologie

DOCUMENT : Texte, Image ou Epure représenté par l'icone < > . Un document est soit une source, stockée physiquement sur le serveur, soit un fragment, portion d'une source.

RESEAU DE DESCRIPTION : Décrit le classement et les liens entre les documents d'une bibliothèque. Il est composé d'arcs (Spécialisations) et de nœuds (Descripteurs).

NOTE : Apporte une précision par rapport à un corpus de documents. Il est représenté par l'icone < >. Son contenu textuel est codé en Unicode : c'est-à-dire qu'il peut contenir entre autres du latin accentué et du grec polytonique.

PARCOURS DE LECTURE : Séquence de documents (en contexte) visités par un utilisateur.

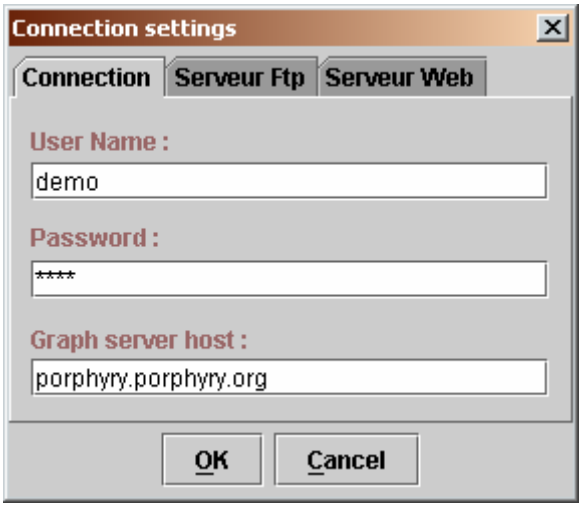
2. Les premiers pas avec Porphyre

Pour lancer le client *Porphyre*, double-cliquez sur le fichier `Client.bat` que vous trouverez à l'aide de l'explorateur dans le sous-dossier `bin` du dossier `porphyry2002`.

2.1. Les paramètres de la connexion

2.1.1. pour le serveur Porphyre

Entrez les paramètres de l'utilisateur pour afficher le bon graphe dans le client *Porphyre*.



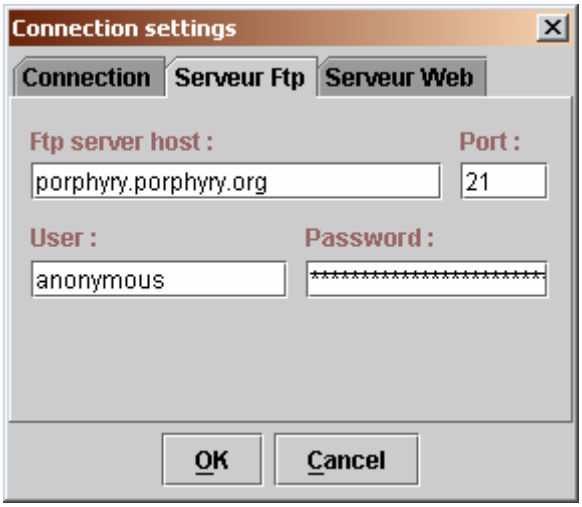
PARAMETRES DE DEMONSTRATION :

User name : `demo`

Password : `demo`

Graph server host : `porphyry.porphyry.org`

2.1.2. Pour le serveur FTP



PARAMETRES DE DEMONSTRATION :

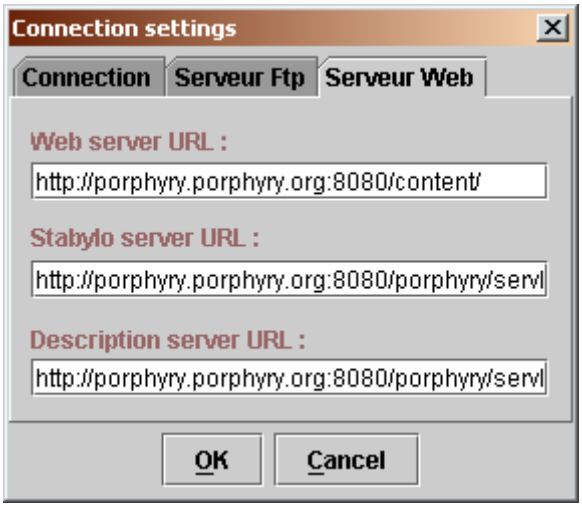
FTP server host : `porphyry.porphyry.org`

Port : `21`

User : `anonymous`

Password : < Votre adresse e-mail >

2.1.3. Pour le serveur Web



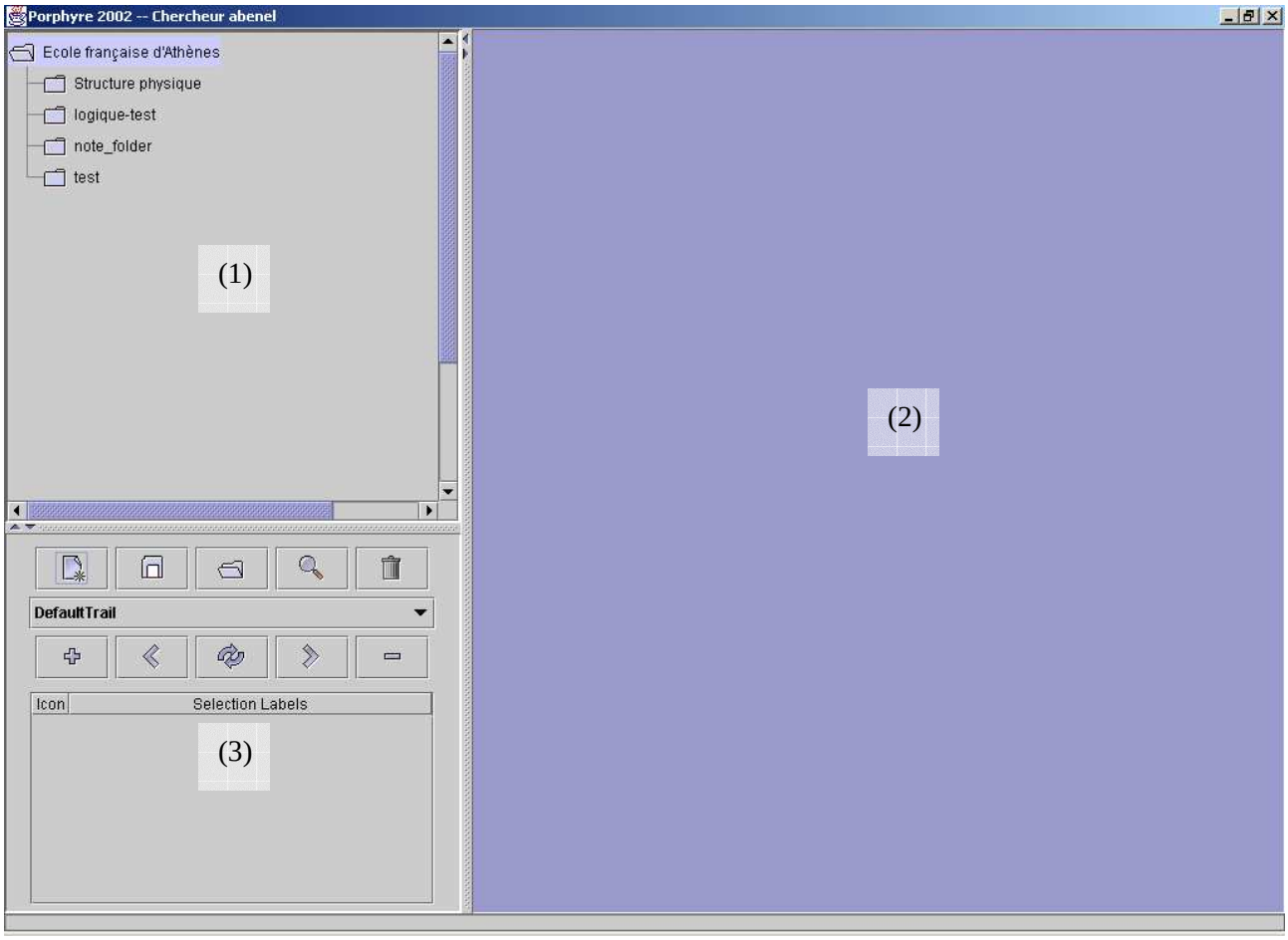
PARAMETRES DE DEMONSTRATION :

Web server URL : `http://porphyry.porphyry.org:8080/content/`

Stabylo server URL : `http://porphyry.porphyry.org:8080/porphyry/servlet/document?doc=`

Description server URL : `http://porphyry.porphyry.org:8080/porphyry/servlet/description?id=`

2.2. Fenêtre de l'environnement du client Porphyre



La fenêtre du client Porphyre comporte trois zones dédiées aux :

- (1) réseaux de description
- (2) corpus de documents
- (3) parcours de lecture

Les dimensions des zones sont modifiable à souhait avec le pointeur de la souris, ou grâce aux symboles et se trouvant sur les limites de chaque zone.

3. Réseau de description

Le réseau affiché est constitué du réseau de l'utilisateur et de ceux des communautés auxquelles il est abonné. L'utilisateur ne pourra modifier que son réseau personnel.

3.1. Navigation

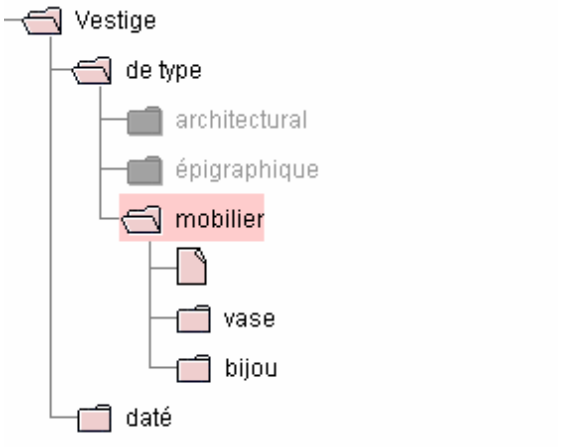
Lors de la sélection d'un descripteur à l'aide d'un clic, celui-ci est rajouté à la requête courante. Une partie du réseau est affichée dans lequel les descripteurs sont associés à un état :

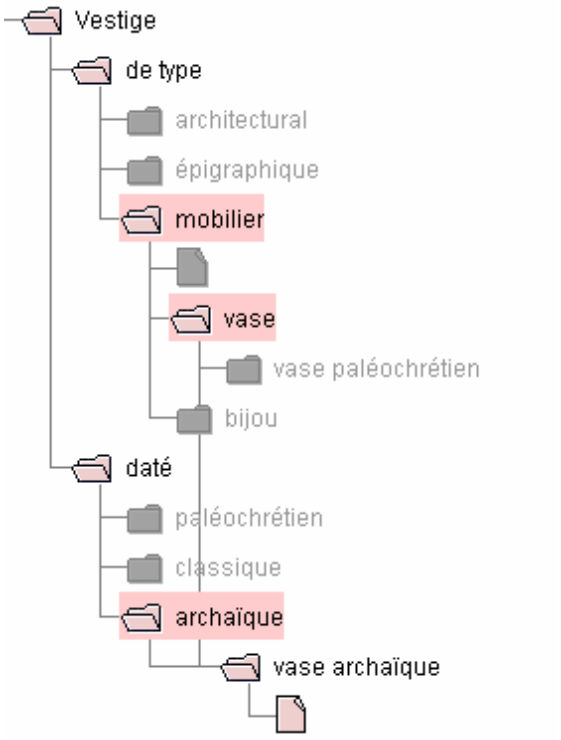
Connu : ce descripteur est associé à l'ensemble du corpus sélectionné. Il fait donc partie de la sélection de manière implicite ou explicite.

Possible : ce descripteur est associé à une partie du corpus sélectionné. Il peut donc être ajouté à la sélection.

Impossible : ce descripteur n'est associé à aucun des documents du corpus sélectionné. Il ne peut donc pas être ajouté à la sélection (on obtiendrait un corpus vide).

Un clic sur un descripteur précédemment sélectionné aura pour effet de le retirer de la requête courante.

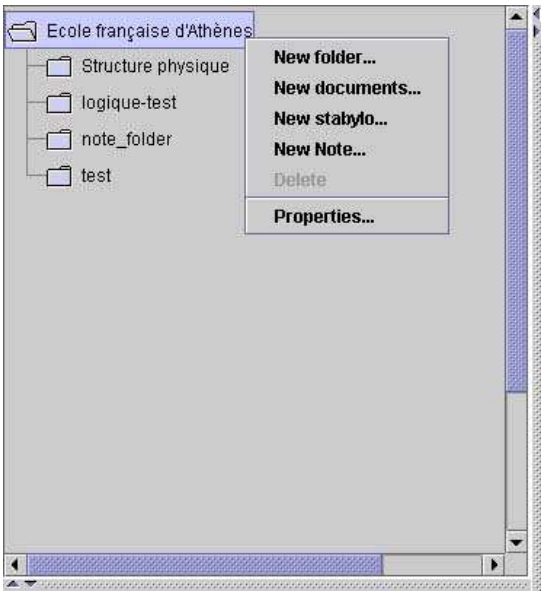




Le réseau n'est pas forcement hiérarchique : un document peut être décrit suivant différentes dimensions d'indexation (facette) et un descripteur lui-même peut appartenir à plusieurs facettes.

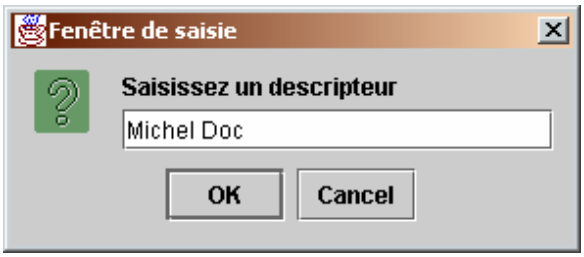
3.2. Fonctionnalités du menu

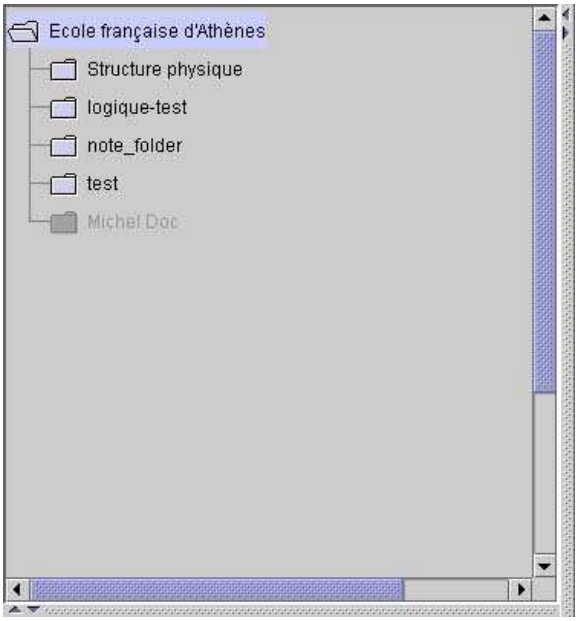
A l'aide du bouton droit de la souris sur un objet du graphe, nous obtenons les fonctionnalités disponibles. Par exemple, ici sur un descripteur "Ecole française d'Athènes".



3.2.1. Créer un dossier

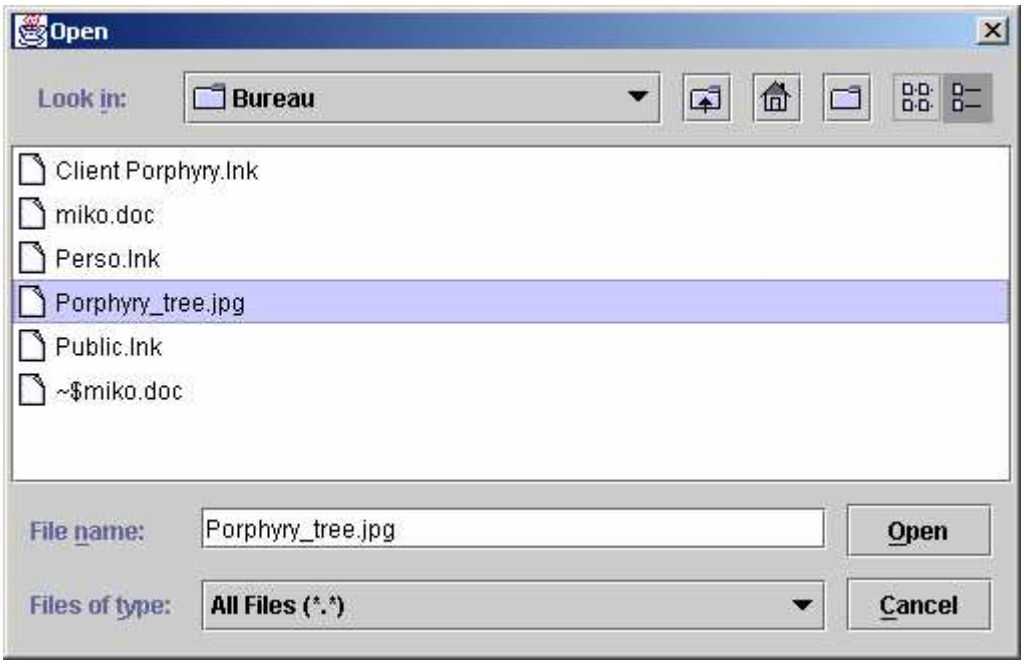
En Sélectionnant la fonction "New folder", nous créons un descripteur dans le réseau à l'endroit indiqué. Il suffit ensuite d'entrer le nom du nouveau descripteur dans la fenêtre de saisie qui apparaît.



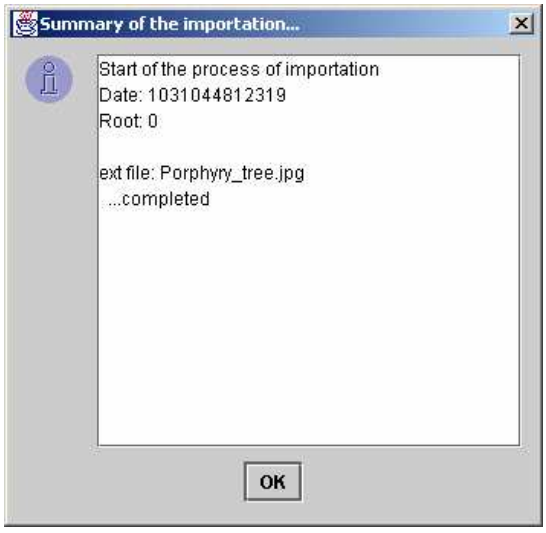


3.2.2. Ajouter un document

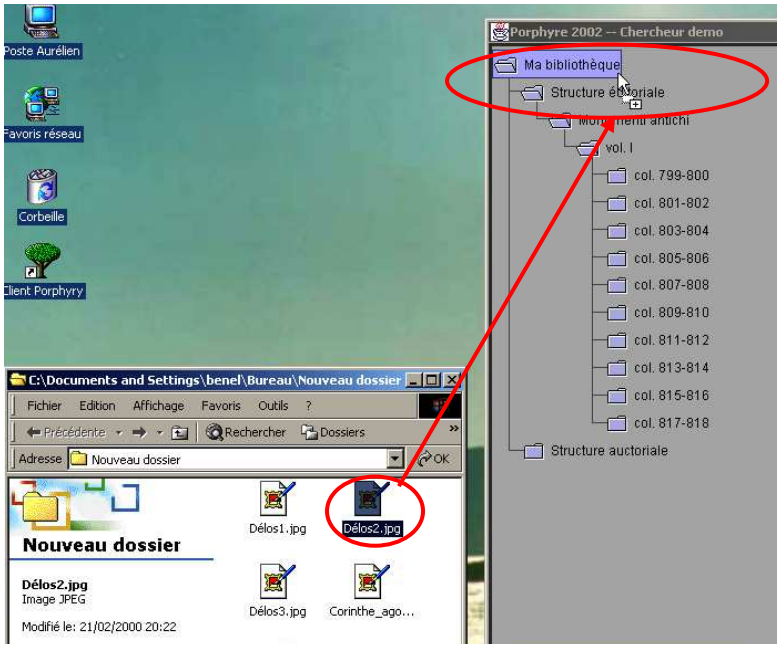
En Sélectionnant la fonction "New documents", nous ajoutons un *document* dans le réseau à l'endroit indiqué. Nous sélectionnons le document correspondant dans la fenêtre d'exploration (à l'heure actuelle seuls les fichiers JPEG sont gérés).



Une fenêtre indique alors le bilan du télédéchargement du document sur le serveur. En cas de succès, le document apparaît dans le réseau.



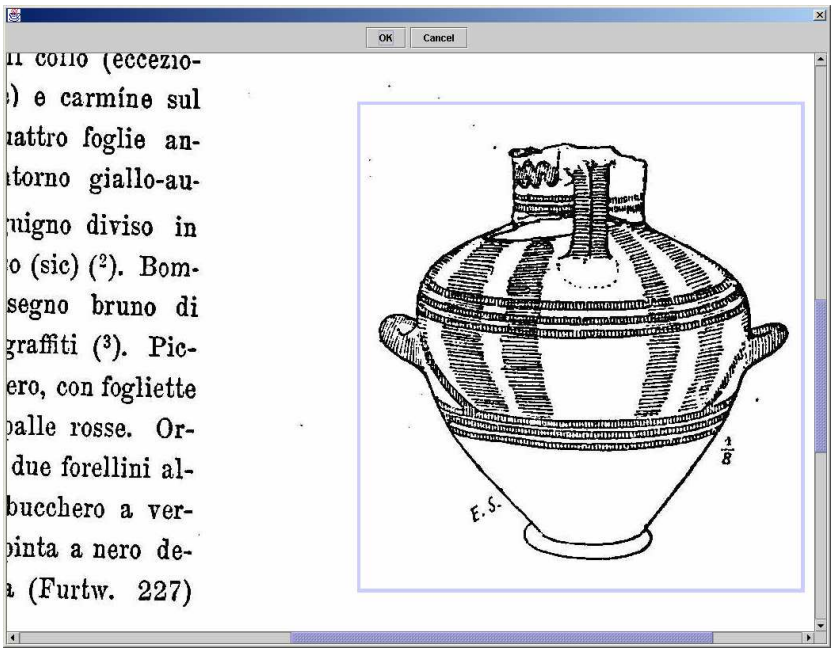
Le même résultat peut être obtenu en faisant glisser le fichier à partir du bureau ou de l'explorateur.



3.2.3. Créer un fragment

En Sélectionnant la fonction "New Stabylo", nous ajoutons un *fragment* du dernier document actif à l'endroit indiqué dans le réseau.

Dans une nouvelle fenêtre apparaît la source à l'échelle 1:1, où l'utilisateur désigne une sélection rectangulaire (à l'aide de la souris). Une fois la sélection validée, le fragment est placé dans le réseau.

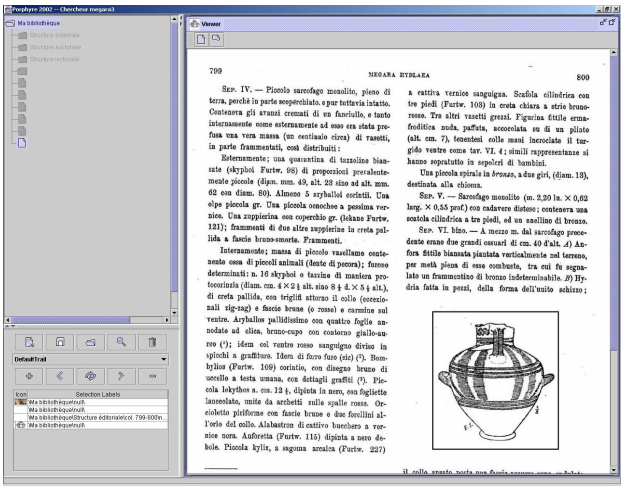


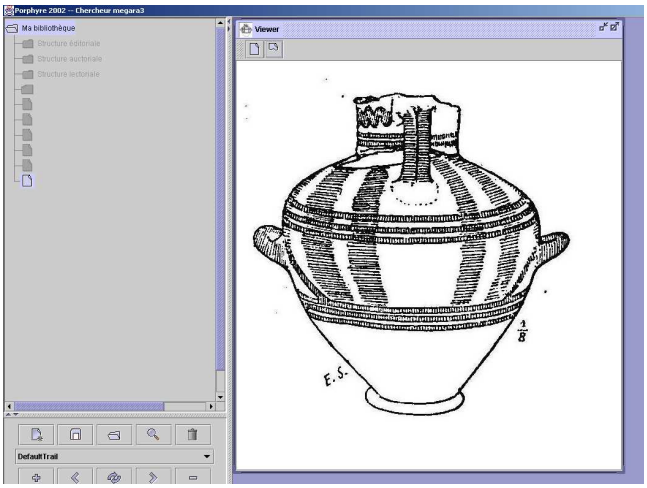
Chaque fragment peut être visualisé suivant deux modes :

en contexte (dans sa source),

hors-contexte (fragment seul).

Dans les deux cas, les fragments, comme les sources, sont redimensionnés en fonction de la largeur de la fenêtre.

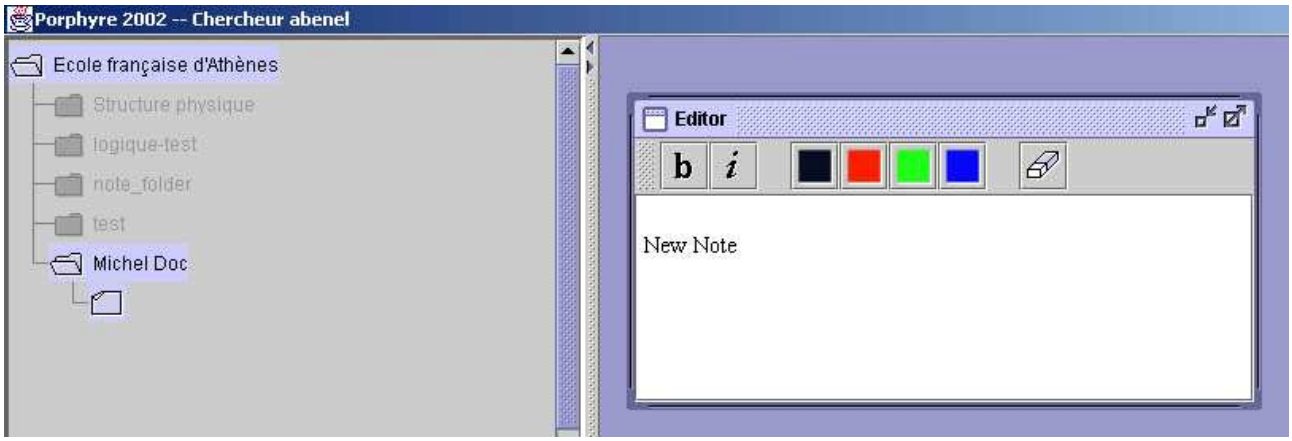




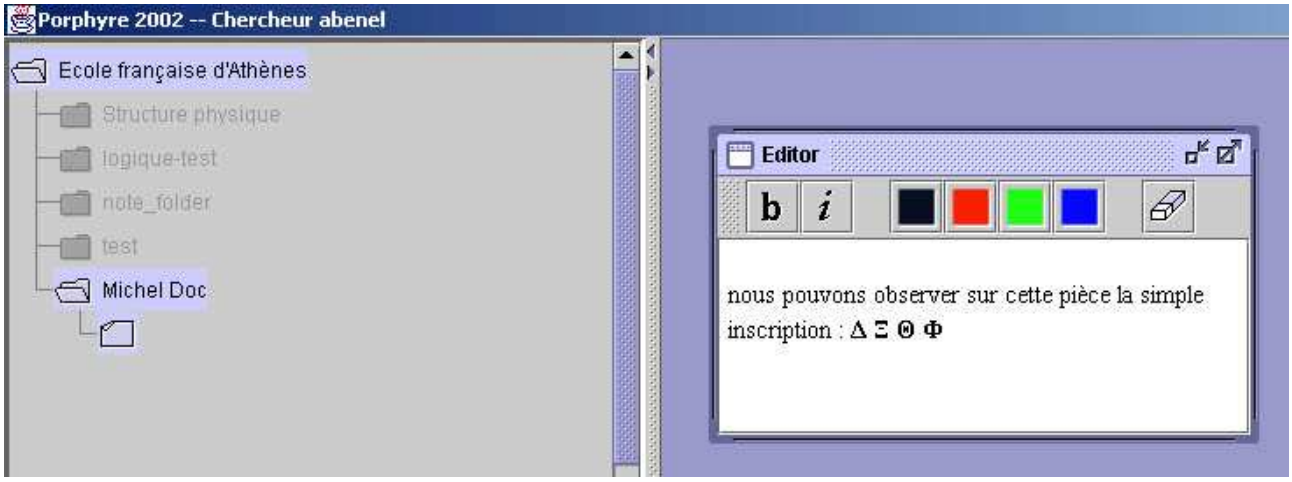
3.2.4. Créer une note

(voir aussi chapitre 4 : "Editeur de notes")

En Sélectionnant la fonction "New Note", nous créons une *Note* dans le réseau à l'endroit indiqué.



Il suffit ensuite de sélectionner la *Note* et de la remplir avec un texte original (Unicode UTF-8, contenant la plupart des alphabets).



Le contenu de la *Note* est sauvegardé automatiquement quand la fenêtre de la *Note* est fermée ou désactivée.

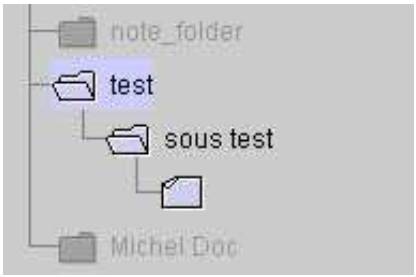
3.2.5. Supprimer un descripteur, un document ou une note

En Sélectionnant la fonction "Delete" sur un objet du réseau, nous supprimons l'objet (*decripteur*, *Document*, *Note*). Attention, s'il s'agit d'un descripteur, il sera supprimé ainsi que ses sous-descripteurs.

3.3. Fonctionnalités supplémentaires

3.3.1. Déplacer

Pour déplacer un objet (*Descripteur*, *Document*, *Note*) au sein du réseau, il suffit de sélectionner l'objet du réseau en maintenant le bouton droit de la souris enfoncé, et de relâcher le bouton à la nouvelle position souhaitée dans le réseau.

<table>
<tr>
<td>-1-
 

Nous voulons déplacer la Note vers cet endroit du réseau.</td>
<td>-2-

Voici l'emplacement actuel de la Note dans le réseau.</td>
<td>-3-

Cliquer sur la Note en maintenant le bouton enfoncé.</td>
</tr>
<tr>
<td>-4-

Déplacer le curseur à l'endroit voulu.</td>
<td colspan="2">-5-

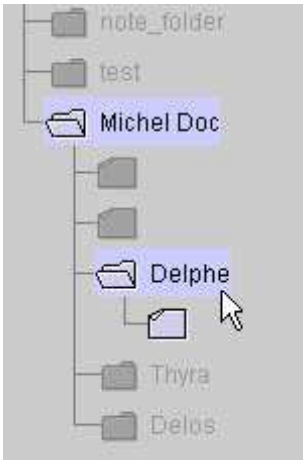
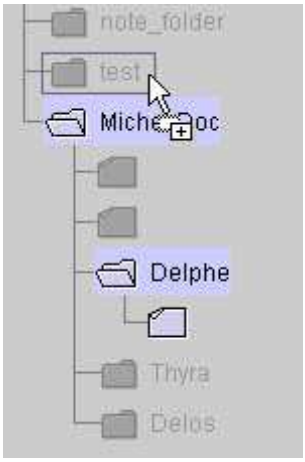
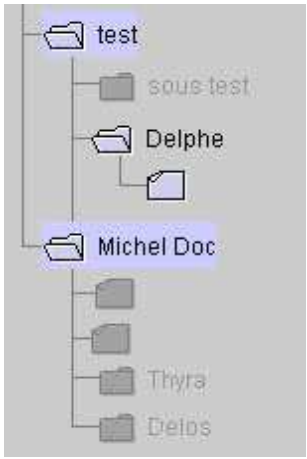
Relâcher le bouton de la souris.
La Note se trouve maintenant à l'endroit voulu dans le réseau.</td>
</tr>
</table>

Astuce ☺ : il est souvent plus aisé d'effectuer le déplacement entre deux clients *porphyre*.

3.3.2. Créer un lien

Nous pouvons référencer un même objet (*Descripteur*, *Document*, *Note*) à des endroits différents du réseau. Nous appèlerons ces spécialisations supplémentaires des liens.

Le principe est le même que pour l'opération "Move" en maintenant la touche "Ctrl" du clavier enfoncée.

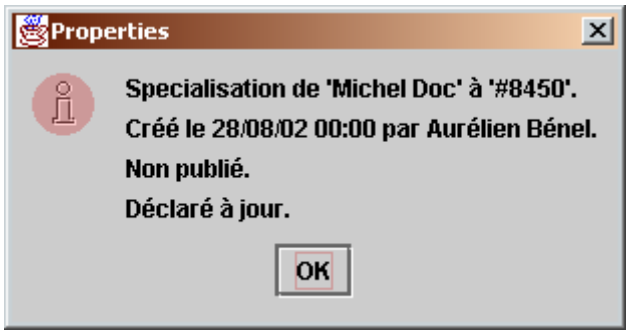
-1-	-2-	-3-
		
Sélection de l'objet à relier à une autre partie du réseau.	Déplacer le curseur vers l'endroit voulu.	- Relâcher le bouton de la souris. - "Delphe" est relié à "Michel Doc" et à "test".

3.3.3. Supprimer un lien

En sélectionnant la fonction "Delete" sur un des liens d'un objet (*Descripteur*, *Document*, *Note*), celui-ci est détruit.

3.3.4. Propriétés

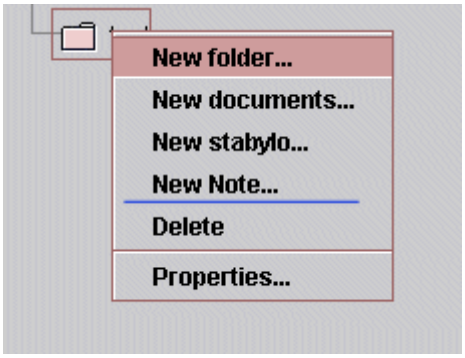
En sélectionnant la fonction "Properties", nous pouvons obtenir des informations sur l'objet du réseau qui est sélectionné.



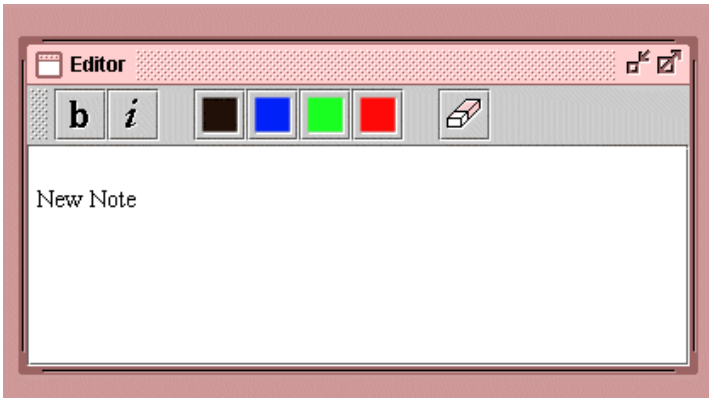
4. Editeur de notes

4.1. Création d'une *Note*

La création d'une note s'effectue depuis le menu lors d'un clic droit sur un descripteur.

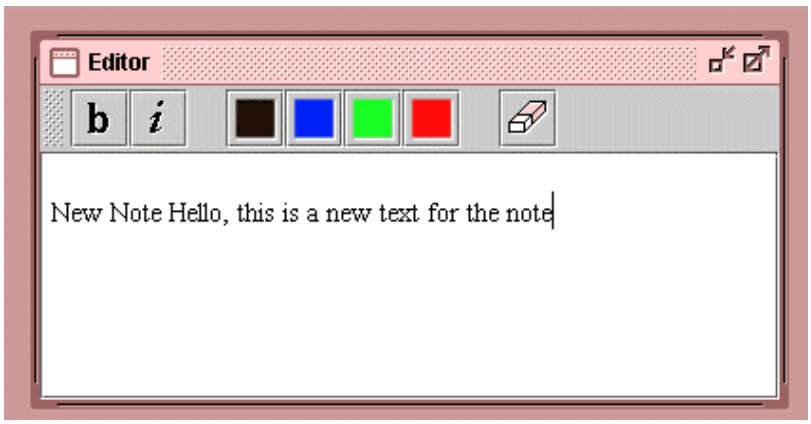


Une Note est alors créée et ajoutée dans le réseau. Cette note est initialisée avec le texte : « New Note ».



4.2. Ajout de texte et sauvegarde de la Note

Lorsque la Note est **active** et le curseur **positionné**, la frappe de touches au clavier entraînera l'ajout du texte correspondant dans la Note à la position du curseur.



Pour enregistrer les changements effectués, il faut désactiver la Note, il suffit pour cela :

Soit de « l'iconifier »

Soit de consulter un autre document

Ou par un clic dans le réseau

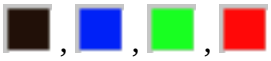
4.3. Mise en Forme du texte

Avant de mettre en forme une zone de texte , il faut la sélectionner à l'aide du clavier ou de la souris.

Ensuite un clic sur le bouton approprié modifie la sélection dans la forme voulue :

En **Gras :**

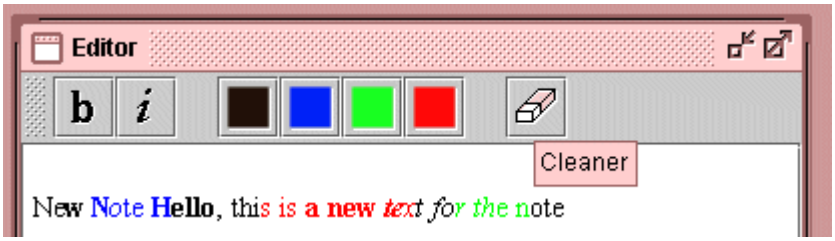
En *Italique :*

En Noir, Rouge, vert ou bleu :  .

Astuce ☺ : une fois la sélection réalisée, il est possible d'y appliquer successivement plusieurs mises en forme.

4.4. Retrait de toutes les Mises en Forme

Il est possible de supprimer toutes les mises en forme du texte de la note. Le bouton « Cleaner » active cette fonctionnalité.

Avant : 

Après :

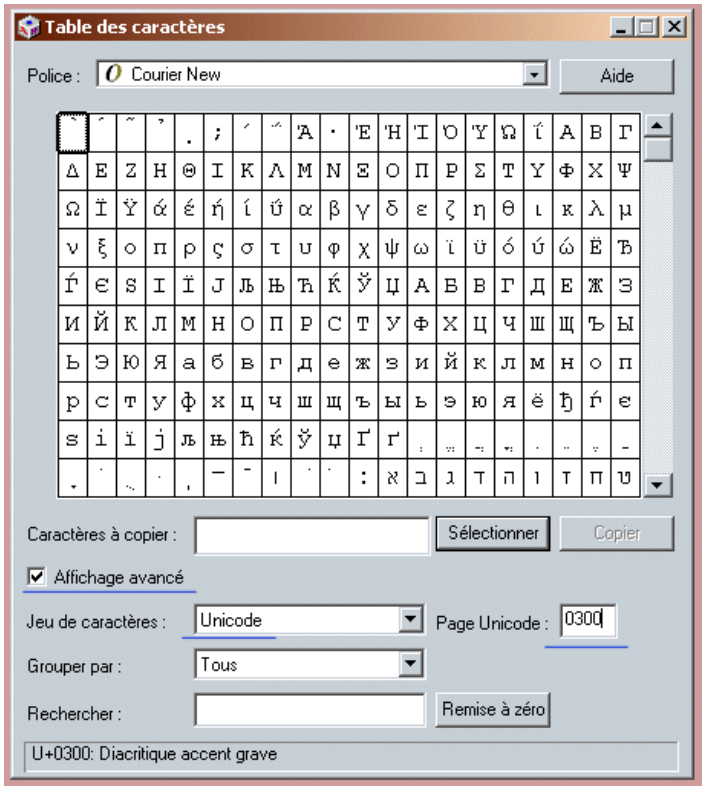
4.5. Ajout de caractères spéciaux

Il est possible d'introduire dans une Note des caractères spéciaux non-saisissables au clavier (ou difficilement).

Pour cela il suffit d'effectuer un « Copier Coller » de ces caractères.

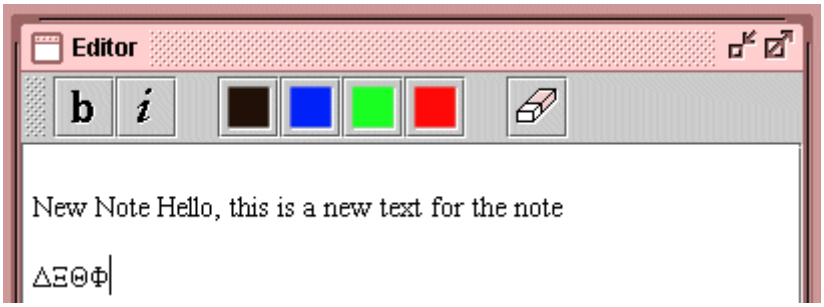
Un moyen assez simple de le faire est d'afficher la *Table des Caractères*. On peut l'obtenir à partir du menu Démarrer sous Windows (le principe est le même avec d'autres systèmes d'exploitation) :

Ensuite, dans une police Unicode (Arial Unicode, par exemple), demander l'affichage des caractères. Les caractères grecs et cyrilliques se trouvent à partir de la position 0300 notamment.



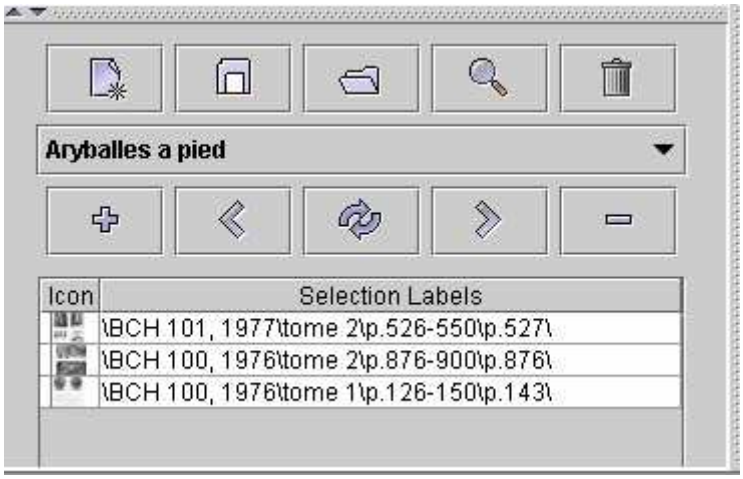
Vous pouvez alors sélectionner les caractères à insérer, et enfin les « Copier ».

De retour dans l'éditeur de Note, il suffit de les Coller <CTRL> + <V>.



5. Parcours de lecture

5.1. Généralité



Le panneau se situant au bas du panneau de contrôle permet la gestion des parcours de lecture de l'utilisateur en cours de connexion.

Par défaut, lorsque l'utilisateur se connecte, un parcours par défaut vide se crée et ce dernier est sélectionné. Pour sélectionner un document dans le réseau et que cette sélection soit prise en compte dans le parcours de lecture, il faut cliquer sur l'icône représentant le document (voir image ci-dessous).

5.2. La gestion des parcours

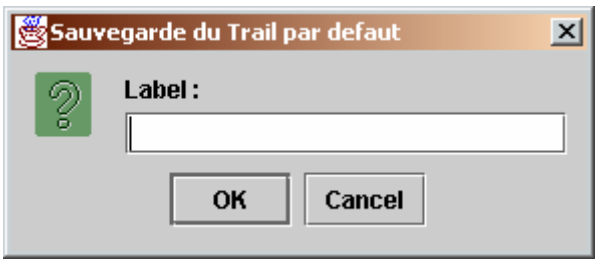
Elle s'effectue par l'intermédiaire de la barre de boutons supérieure,

ainsi qu'avec le « combo box » qui liste les noms des parcours.

Les fonctionnalités suivantes sont permises, par le biais d'un clic sur le bouton correspondant à l'icône :

1) Nouveau parcours par défaut : Cette fonctionnalité permet de remettre le parcours par défaut à vide pendant une session et de placer le « combo box » sur ce parcours.

2) Sauvegarde du parcours par défaut : la sauvegarde est effectuée sur le serveur de données du parcours par défaut avec un label donnée par l'utilisateur par le biais de la fenêtre ci-dessous :



3) Récupération d'une liste de parcours : Si aucun document n'est sélectionné, alors ce sont tous les parcours de l'utilisateur (et des communautés auxquelles il est abonné) qui seront listés dans le « combo box ».

4) Si un document est ouvert, alors cette fonctionnalité permet la récupération de l'ensemble des parcours de l'utilisateur (et de ceux auxquels il est abonné) passant par le même document. Les parcours trouvés sont listés dans la « combo box ».

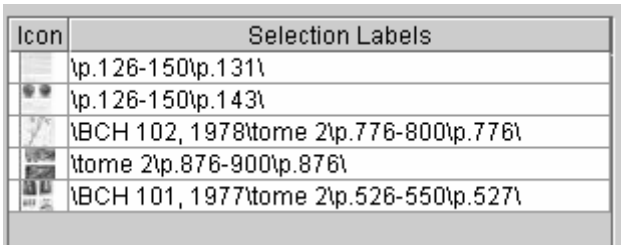
5) Effacement d'un parcours : Cette fonctionnalité permet l'effacement d'un parcours devenu inutile pour l'utilisateur.

6) Chargement d'un parcours : Cette fonctionnalité permet l'affichage des étapes de lecture lors de la sélection d'un parcours dans la « combo box » contenant la liste des parcours : c'est-à-dire, pour chaque étape de lecture, la vignette correspondante et la sélection.

5.3. La gestion des étapes de lecture

Les étapes de lectures sont gérées par la barre de boutons inférieure :

Et par le tableau regroupant la liste d'étapes de lecture :



1) : Ajout à un parcours de l'étape de lecture correspondant au document courant

2) : Retrait de l'étape de lecture sélectionnée

3) : Aller à l'étape de lecture précédente

4) : Aller à l'étape de lecture suivante

5) : Déplacer l'étape de lecture :
- Sélectionnez d'abord une étape en cliquant sur une ligne du tableau
- Cliquez sur le bouton « Déplacer Etape »
- Indiquez la position voulue pour l'étape sélectionnée (commence à 1)

FOLIO ADMINISTRATIF

THÈSE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES

NOM : Bénel PRÉNOM : Aurélien	DATE DE SOUTENANCE : 12 décembre 2003

TITRE : Consultation assistée par ordinateur de la documentation en Sciences Humaines :
Considérations épistémologiques, solutions opératoires et applications à l'archéologie

NATURE : Doctorat — NUMÉRO D'ORDRE : 03 ISAL 0068

ECOLE DOCTORALE : Informatique et Information pour la Société

SPÉCIALITÉ : Documents Multimédia, Images et Systèmes d'Information Communicants

COTE B.I.U. - LYON : T 50/210/19 / ET BIS — CLASSE :

RÉSUMÉ :

Dans les Sciences Humaines, comme dans bon nombre de domaines d'expertise (Médecine, Ingénierie, Droit, Physique…), le document constitue le principal outil de travail et, à ce jour, ni les bases de données ni les systèmes experts n'ont réussi à le supplanter. Dans ces systèmes, en effet, l'usager serait contraint d'insérer ses données dans un modèle fixe et consensuel (schéma de la base, ontologie). Or, dans les domaines d'expertise, chacun apporte son modèle personnel vis-à-vis d'un corpus documentaire de référence. De plus, la confrontation des points de vue stimule une évolution constante des modèles.

Né d'une dynamique interdisciplinaire (Archéologie, Informatique…), le projet Porphyre se propose d'instrumenter ce travail d'expertise. Il est fondé sur l'idée qu'une telle pratique « laisse des traces » : sélections de fragments, documents ajoutés au corpus, organisation de la bibliographie, notes de lecture, etc. A condition que ces traces soient « captées », leur auteur peut désormais les utiliser pour accéder au corpus, ainsi que les partager avec d'autres experts. En somme, puisque le corpus se structure de pair avec le savoir de l'expert au fil de ses changements de problématique et de point de vue, nous proposons un atelier de manipulation de corpus documentaires afin d'assister un travail humain de construction de sens.

MOTS-CLÉS : Bibliothèque numérique, Assistance à l'interprétation, Collaboration, Annotation, Hypermédia.

LABORATOIRES DE RECHERCHE :

École française d'Athènes [EFA], Laboratoire d'InfoRmatique en Images et Systèmes d'information [LIRIS].

DIRECTEURS DE THÈSE: Jean-Marie Pinon, Andréa Iacovella, Sylvie Calabretto.

PRÉSIDENT DE JURY : Roland Etienne

COMPOSITION DU JURY :

Roland Etienne	Professeur, Université Paris I	Président
Jocelyne Nanard	Professeur, Université Montpellier II	Rapporteur
François Rastier	Directeur de Recherche, INaLF (CNRS) Paris	Rapporteur
Jean-Marie Pinon	Professeur, INSA de Lyon	Directeur
Andrea Iacovella	Ingénieur de Recherche, Ecole française d'Athènes	Co-directeur
Sylvie Calabretto	Habilitée à Diriger les Recherches, INSA de Lyon	Co-directrice